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 trigger 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

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 and 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 (Langdale and Kidner 1994, Hall *et al.* 1998). Subsequent characterization revealed that *ZmG2* 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 (Rossini *et al.* 2001, Fitter *et al.* 2002, 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 signalling, defence responses and fruit ripening (Kobayashi *et al.* 2012, Nguyen *et al.* 2014, Li *et al.* 2020). 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 events that led to the evolution of C_4 photosynthesis within the Poales group (Christin and Osborne 2013). In species such as rice that perform C_3 photosynthesis, the *GLK* paralogs are both expressed in mesophyll cells and are functionally redundant as in Arabidopsis (Waters *et al.* 2009, Wang *et al.* 2013). 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 (Woo *et al.* 1970, Sage *et al.* 2011). In the C_4 species 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 such that *GLK2* transcripts accumulate to detectable levels
85 in mesophyll cells of setaria but not maize (Li *et al.* 2010, Chang *et al.* 2012, Wang *et al.* 2013, John *et*
86 *al.* 2014, Tausta *et al.* 2014). Importantly, localization of *ZmG2* transcripts in bundle sheath cells of
87 maize and defects in bundle sheath but not mesophyll chloroplast development in *g2* mutants
88 (Langdale and Kidner 1994, Cribb *et al.* 2001, Rossini *et al.* 2001) demonstrated that *GLK* paralogs in
89 C_4 grasses 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 *ZmGLK1* is expressed in mesophyll cells of maize, and *ZmG2* is expressed and functions specifically
94 in bundle sheath cells, it has been hypothesized that *ZmGLK1* controls chloroplast biogenesis in
95 mesophyll cells (Tausta *et al.* 2014). Here we test this hypothesis by generating and characterizing loss-
96 of-function mutants in maize and setaria, two species that represent independent origins of C_4. Our
97 findings reveal that *ZmG2* and *SvGLK2* orthologs function in the cell-type(s) where the corresponding
98 transcripts accumulate but contrary to expectations, neither *ZmGLK1* nor *SvGLK1* are necessary for
99 chloroplast development in mesophyll cells.
100
101 **RESULTS**
102 ***ZmGLK1* is not necessary for chloroplast biogenesis in mesophyll cells of maize**
103 To determine whether *ZmGLK1* regulates chloroplast biogenesis in maize, loss-of-function mutants
104 were generated using CRISPR/Cas9 (Figure 1A). Four independent mutant lines (referred to
105 collectively as *Zmglk1*) were obtained, all of which exhibited the same overall growth phenotype as
106 segregating wild-type siblings (Figure 1B, C; Figure S1). Ultrastructure analysis revealed that
107 chloroplasts in *Zmglk1* lines were similar to wild-type, with well-developed granal thylakoids in
108 mesophyll cells and agranal thylakoids in bundle sheath cells (Figure 1D-I; Figure S1). Total chlorophyll
109 levels were also similar in wild-type and *Zmglk1* mutants (Figure 1J). As such, we concluded that
110 *ZmGLK1* is not necessary for thylakoid assembly or chlorophyll biosynthesis in bundle sheath or
111 mesophyll chloroplasts of maize.
112
113 An evaluation of chloroplast function in *Zmglk1* mutants was made by measuring photosystem II (PSII)
114 parameters. PSII is a large protein complex comprised of the antenna and chlorophyll molecules that
115 harvest light and together initiate the electron transfer reactions of photosynthesis (Nickelsen and
116 Rengstl 2013). In maize, PSII accumulates in the granal thylakoids of mesophyll cells and is largely
117 absent from the agranal bundle sheath cell chloroplasts (Woo *et al.* 1970). As such, PSII function is a
118 proxy for mesophyll cell chloroplast function. Fluorescence measurements were therefore taken in light
119 and dark-adapted plants to quantify PSII operating efficiency (PhiPSII) and maximum quantum

120 efficiency of PSII photochemistry (Fv/Fm), in both wild-type and *Zmglk1* mutant plants. Figures 1K-L
121 show that PSII is operating normally in *Zmglk1* mutants, suggesting that *ZmGLK1* is not necessary for
122 chloroplast biogenesis or function in maize mesophyll cells.
123
124 ***SvGLK1* is not necessary for chloroplast biogenesis in mesophyll cells of setaria**
125 Maize and setaria have different origins of C_4 photosynthesis within the PACMAD clade (Li and Brutnell
126 2011) but like the orthologous *ZmGLK1*, *SvGLK1* transcripts accumulate preferentially in mesophyll
127 cells (John *et al.* 2014). Phylogenetic analysis and amino acid sequence comparisons revealed that
128 *SvGLK1* shares many structural features with *ZmGLK1* (Figure S2). To assess whether *SvGLK1* plays
129 a role in chloroplast biogenesis, two independent loss-of-function mutants were generated in setaria
130 using CRISPR/Cas9 (Figure 2A). As in maize, *Svglk1* mutants exhibited overall plant growth (Figure 2B-
131 D) and chloroplast ultrastructure (Figure 2E-J) phenotypes that were similar to segregating wild-type
132 siblings, with thylakoid membranes developing normally in both mesophyll and bundle sheath cells
133 (Figure 2E-L). Furthermore, there was no significant difference in chlorophyll levels between *Svglk1*
134 mutants and wild-type (Figure 2M), and electron flow through PSII was unperturbed (Figure 2N-O).
135 Collectively, these results suggest that the preferential expression of *GLK1* genes in mesophyll cells of
136 C_4 grass species does not reflect a requirement for gene function during chloroplast development
137 and/or maintenance.
138
139 ***SvGLK2* regulates chloroplast development in both mesophyll and bundle sheath cells of setaria**
140 Given that dimorphic chloroplast development is observed in bundle sheath and mesophyll cells of
141 both maize and setaria, we next sought to determine whether *SvGLK2,* like its ortholog *ZmG2* (Figure
142 S2), regulates bundle sheath chloroplast development. To this end, CRISPR/Cas9 was used to generate
143 loss-of-function *Svglk2* mutants (Figure 3A). In contrast to maize where *g2* mutants are weaker than
144 wild-type but are viable and fertile, any *Svglk2* mutant plants that regenerated after transformation did
145 not progress beyond the seedling stage before dying. The only viable line recovered from tissue culture
146 was heterozygous for *SvGLK2* and this was used to isolate a null homozygous line via self-pollination.
147 *Svglk2* 8G5 mutants, henceforth referred to as *Svglk2*, exhibited severe growth defects, had pale yellow
148 leaves and failed to complete the life cycle (Figure 3B). Leaf cross sections showed misshapen and
149 smaller chloroplasts in the bundle sheath cells of *Svglk2* mutants compared to wild-type (Figure 3C-D),
150 and observations of chloroplast ultrastructure revealed impaired chloroplast anatomy with
151 discontinuous thylakoids that occupied less of the organelle than in wild-type (Figure 3E-F), albeit a
152 reduction that was not statistically significant (Figure 3I). More striking, given the bundle sheath specific
153 perturbations observed in maize *g2* mutants, was the aberrant phenotype of mesophyll chloroplasts in
154 *Svglk2* mutants. Compared to wild-type there was a significant reduction in thylakoid membrane
155 coverage (15% less than wild-type) and a noticeable reduction in the size and frequency of grana
156 (Figure 3G-H, J). As in bundle sheath cells, thylakoid membranes in mutant mesophyll chloroplasts

157 appeared fragmented (Figure 3H). The observed perturbations in bundle sheath and mesophyll cells
158 were accompanied by a reduction in leaf chlorophyll content of more than 80% (Figure 3K). These data
159 reveal a direct association between the presence of *SvGLK2* and *ZmG2* transcripts in a particular cell-
160 type and a requirement for gene function during chloroplast development, with the presence of *SvGLK2*
161 transcripts in mesophyll cells of setaria (albeit at lower levels than in bundle sheath cells) (John *et al.*
162 2014) reflecting a role in chloroplast biogenesis in both cell-types.
163
164 ***GLK2* transcript levels are not upregulated in *glk1* mutants of maize or setaria**
165 Given the direct correspondence between *ZmG2* and *SvGLK2* transcript accumulation profiles and
166 necessity of gene function for chloroplast development, the absence of such correspondence for
167 *ZmGLK1* and *SvGLK1* orthologs is intriguing. It is particularly notable because all *GLK* genes
168 characterized to date have a role in chloroplast biogenesis, albeit a redundant role alongside paralogs
169 in many cases (Fitter *et al.* 2002, Wang *et al.* 2013). Although *SvGLK2* may compensate for loss of
170 *SvGLK1* function because *SvGLK2* transcripts comprise ~40% of total *GLK* transcript levels in
171 mesophyll cells of wild-type setaria, this is unlikely to be the case in maize where *ZmG2* transcripts only
172 comprise ~6% of total *GLK* transcript levels in mesophyll cells of wild-type plants (Figure S3). However,
173 loss of *ZmGLK1* function could lead to altered regulation of *ZmG2* and/or to the upregulation of genes
174 encoding the GATA transcription factors GNC and CGA1 which also regulate chloroplast development
175 (Chiang *et al.* 2012, Hudson *et al.* 2013) and have been shown to partially rescue loss of GLK function
176 in Arabidopsis (Zubo *et al.* 2018). To test this hypothesis, transcript levels of *ZmGLK1*, *ZmG2*, *ZmGNC*
177 and *ZmCGA1* were quantified in maize using the *Zmglk1* gene edited mutants generated above
178 alongside the previously reported transposon-induced *Zmg2-bsd1-s1* mutant (Cribb *et al.* 2001). In
179 each case the mutant allele produces transcripts that encode non-functional protein(s). Figure 4 reveals
180 that loss of *Zmglk1* function does not lead to consistently elevated levels of *ZmGLK1* (Figure 4A), *ZmG2*
181 (Figure 4B), *ZmGNC* (Figure 4C) or *ZmCGA1* (Figure 4D) transcripts. A small but significant increase
182 in *ZmG2*, *ZmGNC* and *ZmCGA1* transcript levels was seen in one of the four *Zmglk1* mutant lines but
183 that cannot explain how chloroplasts develop normally in all four lines. We thus conclude that
184 transcription of *ZmG2, ZmGNC* and *ZmCGA1* is not upregulated in mesophyll cells when *Zmglk1*
185 function is lost and thus that, unlike in all other contexts examined, chloroplasts in the mesophyll cells
186 of maize can develop normally in the absence of any GLK activity within that cell-type.
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. In the case of

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

230 and bundle sheath cells, and those expressing *ZmG2* accumulate transcripts encoding functional
231 ZmG2 and SvGLK1 proteins, similarly in both cell types.
232
233 To determine whether *ZmGLK1* and/or *ZmG2* can rescue chloroplast ultrastructure defects in
234 mesophyll and/or bundle sheath cells of *Svglk2* mutants, confocal images and transmission electron
235 micrographs of transverse leaf sections from all six transgenic lines were examined. In all cases,
236 chloroplasts were restored to normal size (Figure 6A-C, Figure S4), with predominantly agranal
237 thylakoids in bundle sheath cell chloroplasts (Figure 6D-F, Figure S4) and granal thylakoids in
238 mesophyll chloroplasts (Figure 6G-I, Figure S4). Quantification revealed that with the exception of the
239 one line that had low levels of transgene expression (#17808-15254), thylakoid occupancy levels in
240 mesophyll (Figure 6J) and bundle sheath (Figure 6K) chloroplasts of transgenic lines were statistically
241 equivalent (or in one case higher) than in wild-type plants. Consistent with ectopic expression of *GLK*
242 genes, mature chloroplasts were also observed in the leaf vasculature of all complemented lines, more
243 with *ZmG2* than *ZmGLK1* (Figure 6A-C, Figure S4). Irrespective of which maize ortholog was
244 overexpressed, the total chlorophyll content in leaves of transgenic plants was restored to wild-type
245 levels (Figure 6L). Notwithstanding the fact that SvGLK1 may be functional in bundle sheath and
246 mesophyll cells of all transgenic plants, collectively these data demonstrate that ZmGLK1 and ZmG2
247 are functionally equivalent in the context of the C_4 setaria leaf and that the dimorphic identity of bundle
248 sheath and mesophyll cell chloroplasts develops regardless of which maize ortholog is overexpressed.
249
250 **DISCUSSION**
251 Prior to this study it was known that *GLK* genes are duplicated in many land plant species and that in
252 species that carry out C_3 photosynthesis the duplicate genes act redundantly to regulate chloroplast
253 biogenesis in all photosynthetic cells (Waters *et al.* 2008, Bravo-Garcia *et al.* 2009, Wang *et al.* 2013).
254 It had also been hypothesized that *GLK1* and *GLK2* genes regulate dimorphic chloroplast development
255 in the two photosynthetic cell-types of C_4 species, with each gene functioning in one of the two cell-
256 types (Rossini *et al.* 2001). This hypothesis was based on the observation that *GLK1* genes are
257 preferentially expressed in mesophyll cells of maize, sorghum and setaria whereas *GLK2/G2* genes are
258 preferentially expressed in bundle sheath cells (Li *et al.* 2010, Wang *et al.* 2013, John *et al.* 2014), and
259 loss of *Zmg2* function in maize perturbs bundle sheath but not mesophyll chloroplast development
260 (Cribb *et al.* 2001). Through the phenotypic characterization of *glk1* loss-of-function mutants in maize
261 and setaria (Figures 1, 2 & 4), *glk2* loss-of-function mutants in setaria (Figures 3 & 4), and *glk2* setaria
262 mutants complemented with maize *GLK1* and *GLK2* genes (Figure 5 & 6), we have revealed a more
263 complex situation both with respect to the developmental processes operating within individual C_4
264 species and to likely evolutionary trajectories.
265
266

267 **GLK paralogs have equivalent functional capacity**

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

291

292 Given that *GLK* gene paralogs encode functionally equivalent proteins both in C_3 species where they
293 act redundantly and in the C_4 species maize, the question arises of why the paralogs are expressed in
294 a cell-type preferential manner in C_4 grass species (at least in the ones that have been examined).
295 Estimates based on the topology of BEP (C_3 species only) and PACMAD (C_3 & C_4 species) clades in
296 the Poales suggest that maize diverged from rice ~80 million years ago (Gallaher *et al.* 2022) and that
297 within the PACMAD clade, maize and sorghum share a common history of C_4 evolution whereas
298 independent events occurred in maize and setaria around ~45 million years apart (Gallaher *et al.* 2022).
299 Both *GLK* gene paralogs are predominantly expressed in mesophyll cells of C_3 rice leaves, with *GLK1*
300 transcripts accumulating at higher levels than *GLK2* (Hua *et al.* 2021), and gene function being
301 redundant (Wang *et al.* 2013). Notably, both transcripts can also be detected in bundle sheath cells, at
302 levels roughly equivalent to those of *GLK2* in mesophyll cells, suggesting that the few chloroplasts that
303 develop in rice bundle sheath cells are also regulated by the redundant action of both genes. By

304 contrast, an element of sub-functionalization is apparent in both maize and setaria, with the *GLK2*
305 paralog expressed preferentially in bundle sheath cells, almost exclusively so in maize. Given that the
306 evolution of C_4 required activation of photosynthesis and a significant increase in chloroplast volume in
307 bundle sheath cells, selection on the promoter of the *GLK2* ortholog to enable high levels of expression
308 in that cell type may have been a critical step. Indeed, whether due to selection or drift, a phylogenetic
309 signal of variation is evident between C_3 and C_4 species in the *GLK2* but not the *GLK1* promoter
310 sequence (Figure S3). In this scenario, older C_4 lineages in the Poales would display more specific
311 localization of *GLK2* transcripts in bundle sheath cells than younger lineages, a situation that is evident
312 in comparisons between maize and setaria (Figure S3). However data from a number of different C_3,
313 C_4 and intermediate species (Washburn *et al.* 2021) (Figure S3) invoke a more complex scenario
314 because preferential expression of *GLK2* paralogs in bundle sheath cells is evident in species where
315 decarboxylation reactions occur in the bundle sheath chloroplasts (NADP-ME type) but not in species
316 where decarboxylation occurs in the mitochondria (NAD-ME type) or in the cytoplasm (PEP-CK type).
317 This suggests that preferential localization of GLK2 in bundle sheath cells is associated with specific
318 aspects of chloroplast function as opposed to general activation of photosynthesis.
319
320 **Despite equivalent functional capacity, GLK2 plays a dominant role in both maize and setaria**
321 The difference in levels of bundle sheath specificity of *GLK2* transcripts in maize versus setaria can
322 partially explain the different phenotypes observed in loss-of-function mutants. In maize, where *ZmG2*
323 is almost exclusively expressed in bundle sheath cells, *Zmg2* mutants exhibit perturbed chloroplast
324 development only in that cell-type. By contrast, expression of *GLK2* in both mesophyll and bundle
325 sheath cells of setaria is associated with defective chloroplast development in both cell-types (Figure
326 3). This observation was initially counter-intuitive because it would be expected that, as in C_3 plants,
327 the high levels of *SvGLK1* transcripts in mesophyll cells would compensate for loss of SvGLK2 function.
328 However, we observed that *SvGLK1* transcript accumulation was virtually abolished in *Svglk2* mutants
329 and levels of the non-functional *Svglk2* transcripts were also reduced (Figure 4). As such, we conclude
330 that SvGLK2 activates expression (or transcript stabilization) both of itself and of *SvGLK1,* and that the
331 *Svglk2* mutant phenotype reflects loss of function of both genes. Collectively, these data suggest that
332 SvGLK2 is the primary regulator of chloroplast development in both bundle sheath and mesophyll cells
333 of setaria.
334
335 Whereas normal mesophyll chloroplast development in the absence of GLK1 function can be explained
336 in setaria because SvGLK2 likely compensates for loss of GLK1 function, the mutant phenotype in
337 maize was unexpected. In every land plant species examined to date, a role for *GLK* genes in
338 chloroplast development has been revealed, albeit often a redundant role with paralogs operating in
339 the same cell-type. Because ZmGLK1 is highly expressed in maize mesophyll cells and no other GLK-
340 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 in the bundle sheath (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.

METHODS

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 µmol photon m^{-2} s^{-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 µmol photon m^{-2} s^{-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 and Haeussler 2018) online tool (http://crispor.tefor.net/) to achieve high editing efficiency for the target genes *ZmGLK1* (GRMZM2G026833), *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 µM of each with 1x Ligase Buffer (New England Biolabs) and 10 U of T4 Polynucleotide Kinase (New

378 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
379 M NaCl was added, and the reaction was placed in a water bath heated to 95 °C for 2 min. The reaction
380 was cooled to room temperature before being diluted 100-fold and cloned downstream of the *TaU6*
381 promoter (module pJG310) using the Golden Gate one-step one-pot protocol (Engler *et al.* 2009) with
382 an appropriate restriction enzyme (*Esp3I*). This module was then combined with another level 1 module
383 containing *TaCas9* downstream of the maize ubiquitin promoter (pJG471) into the final acceptor and
384 binary vector pTC278. Both modules and the vector were kindly provided by Dr Dan Voytas (University
385 of Minnesota).
386
387 To overexpress *ZmGLK1* and *ZmG2*, the coding sequences of the maize genes were domesticated to
388 avoid recognition sites for the type II restriction enzymes used in Golden Gate assembly: *BsaI*, *BpiI*,
389 *Esp3I* and *DraIII*. The final sequence was synthesized as a level 0 module with the appropriate SC
390 flanking sites (Engler and Marillonnet 2013). *ZmGLK1* and *ZmG2* level 0 modules (EC17054 and
391 EC17056) were cloned downstream of the maize ubiquitin promoter ($ZmUBI_{pro}$, EC15455) and
392 upstream of a *nos* terminator (tNOS, EC41421) in level 1 backbone position forward 2 using the Golden
393 Gate one-step one-pot protocol (Engler *et al.* 2009). The previously described *hygromycin*
394 *phosphotransferase* (HygR) coding sequence cloned downstream of the rice actin promoter ($OsACT_{pro}$)
395 was used as the selectable marker (Vlad *et al.* 2019). Level 1 modules were assembled into the binary
396 vector pAGM4723 to obtain the constructs EC17808 and EC17810 (hereafter referred to as 17808 and
397 17810, respectively) depicted in Figure S4A and S4B.
398
399 **Plant transformation**
400 *S. viridis* accessions A10 and ME034V were used to obtain CRISPR/Cas9-driven mutations for *SvGLK1*
401 and *SvGLK2*, respectively. *Agrobacterium tumefaciens* strain AGL1 harbouring the construct of interest
402 was co-cultivated with callus generated from seeds of *S. viridis*. Callus induction, transformation and
403 seedling regeneration were performed as described by VanEck and collaborators (Van Eck and
404 Startwood 2015). Complementation lines were obtained by transforming constructs 17808 and 17810
405 into callus generated from a population of seeds that segregated for the gene-edited *Svglk2* allele but
406 no longer segregated the original t-DNA which had been recombined out. The presence of t-DNA in
407 regenerated seedlings was confirmed via duplex PCR (Table S1; JLF_Ox12 and JLF_Ox13 amplify the
408 HygR gene; JLF_Ox14 and JLF_Ox15 amplify an endogenous gene used as a DNA quality control).
409 The occurrence of gene editing by CRISPR/Cas9 was verified by PCR amplification of the region
410 flanking the expected mutation site (Table S1; JLF_Ox45 and JLF_Ox46 for *SvGLK1*; JLF_Ox53 and
411 JLF_Ox54 for *SvGLK2*) followed by endonuclease digestion with *MmeI* or *Tsp45I* for *SvGLK1* and
412 *SvGLK2*, respectively, following the manufacturers' guidelines.
413

414 *Z. mays* Hi-II was transformed by the Wisconsin Crop Innovation Centre (University of Wisconsin)
415 according to their transformation pipeline. The presence of the t-DNA was assessed in T1 lines by PCR
416 (Table S1; JLF_Ox178 and JLF_Ox179) and *Zmglk1* mutants were identified by PCR amplification of
417 the region flanking the expected mutation site (Table S1; JLF_Ox39 and JLF_Ox175) followed by
418 endonuclease digestion with *HinfI* according to the manufacturer's instructions.
419
420 For all lines, mutated sequences were verified by Sanger sequencing (Source BioScience, UK).
421 Experimental analyses were performed on homozygous plants in the T2 or subsequent generations.
422
423 **Confocal imaging**
424 The mid-portion of fully expanded leaf 7 from 21 DAS *S. viridis* plants was fixed in 4 %
425 paraformaldehyde and 0.2 % glutaraldehyde in 25 mM phosphate buffer pH 7.2. After fixation, leaves
426 were stored in 0.2 M Na-EDTA pH 9.0. Free-hand cross sections were imaged using a Leica SP9
427 confocal microscope with an Argon laser set to excite chlorophyll at 663 nm wavelength and to emit
428 from 650-800 nm.
429
430 **Ultrastructure analysis via transmission electron microscopy**
431 Sample fixation and preparation were performed by the Electron Microscopy Facility team at the Sir
432 Willian Dunn School of Pathology (University of Oxford). Leaf discs of 2 mm diameter were collected
433 from leaf 5 of 40 DAS maize or 25 DAS *S. viridis* plants and placed into 3 % glutaraldehyde in 0.06 M
434 Sorenson's phosphate buffer pH 7.2. Samples were transferred to a Leica AMW microwave tissue
435 processor for fixation, osmification (with 1 % OsO_4), dehydration with an acetone series, and resin
436 infiltration (LRW Hard). The samples were then removed from the microwave and transferred to fresh
437 resin for four days, renewing the resin twice a day. The embedded samples were transferred into gelatin
438 capsules and filled with resin for polymerization at 60 °C for about 65 h. Sections of 150 nm were cut
439 using a Leica UC7 ultramicrotome, collected onto formvar-coated copper slot grids and post-stained
440 with lead citrate. Images were acquired on a Thermo Fischer Tecnai T12 TEM with a Gatan OneView
441 CMOS camera.
442
443 Thylakoid occupancy of chloroplasts was quantified using ImageJ. The thylakoid area was measured
444 by adjusting the image threshold to create a binary mask on which only the thylakoid membrane was
445 apparent. This area value was divided by the chloroplast planar area to give the thylakoid occupancy
446 level.
447
448 **Chlorophyll extraction and quantification**
449 Leaf discs were harvested from fully expanded leaf 5 at 30 DAS for maize and 21 DAS for *S. viridis*,
450 and immediately frozen in liquid nitrogen. For total chlorophyll extraction, finely powdered tissue was

451 resuspended in 80 % acetone buffered in 100 mM Tris-HCl pH 8.0 (Chazaux *et al.* 2022). Following a
452 centrifugation step, the absorbance of each sample was read in technical triplicates in a FLUOstar
453 Omega spectrometer (BMG Labtech) at 646 and 663 nm. Background absorbance was measured at
454 750 nm. Absorbance values recorded at 750 nm were subtracted from values recorded at 646 and 663
455 nm for normalization purposes. Technical replicates were averaged for each sample. Concentration of
456 chlorophylls a + b were calculated in micrograms (µg) per leaf area using the following the equation:
457 (19.54 x Abs_{646} + 8.29 x Abs_{663}) x dilution factor (Porra *et al.* 1989).
458
459 **Fluorescence quantification of PSII**
460 Photosystem II (PSII) fluorescence was measured using a LiCOR Li-6800 portable photosynthesis
461 system. LiCOR chamber conditions were set to a flow intensity of 500 umol s^{-1}, delta pressure of 0.2
462 kPa, relative air humidity of 50 %, reference concentration of CO_2 at 400 umol mol^{-1} and temperature
463 exchange at 25 °C. For quantification of PSII operating efficiency (PhiPSII), plants were adapted to light
464 under the conditions described above for at least 1.5 h and actinic light intensity on the LiCOR chamber
465 was set to 1000 µmol photon m^{-2} s^{-1}. After clamping, each leaf was adapted for a few minutes until
466 stabilization was reached before logging the fluorescence value. Quantification of maximum quantum
467 efficiency of PSII photochemistry (Fv/Fm) was performed on plants that were dark-adapted for at least
468 16 h. LiCOR actinic lights were off. In both cases, the fluorescence action log was set to FoFm (dark),
469 FsFm' (light), rectangular flash type with red target at 8000 µmol photon m^{-2} s^{-1}, duration of 1000 ms,
470 output rate of 100 Hz and margin of 5 points.
471
472 **Gene expression analysis via qPCR**
473 The mid-portion of fully expanded leaf 5 from at least three biological replicates of 30 DAS maize or 21
474 DAS *S. viridis* were harvested and immediately frozen in liquid nitrogen. Tissue was ground prior to
475 total RNA extraction using a RNeasy Plant Mini Kit (Qiagen) according to the manufacturer's
476 instructions. To avoid any DNA contamination, samples were treated with TURBO DNA-free DNase
477 (Thermo Fischer Scientific) according to the manufacturer's guidelines. For first strand cDNA synthesis,
478 100 ng of total RNA was added to 1 µl of Maxima Enzyme Mix (Thermo Fischer Scientific) and 1x
479 Reaction Mix to a final volume of 10 µL. The mixture was incubated at 25 °C for 10 min, followed by 50
480 °C for 15 min. Enzyme inactivation was carried out at 85 °C for 5 min.
481
482 Amplification reactions were performed in a StepOnePlus Real-Time PCR System (Thermo Fischer
483 Scientific) using SYBRGreen to monitor dsDNA synthesis. The reaction mixtures contained 5 µL of
484 diluted cDNA (1:25), 0.2 µM of each primer and 10 µL of SYBR™ Green PCR Master Mix (Thermo
485 Fischer Scientific) in a total volume of 10 µL. The reaction cycles began with a 5 min denaturation step
486 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
487 cycle, the fluorescence was measured at 60 °C for 35 s. The melting curve was produced by a cycle of

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 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 and 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 and 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.

LIST OF SUPPLEMENTARY MATERIAL

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

Figure S1. Phenotypic characterization of three independent *Zmglk1* mutant lines.

Figure S2. Comparison of *GLK* sequences.

Figure S3. *GLK* transcript accumulation in bundle sheath and mesophyll cells.

Figure S4. Phenotypic characterization of two independent lines of *Svglk2* complemented with *ZmGLK1* or *ZmG2*.

AUTHOR CONTRIBUTIONS

J.L-F and J.A.L conceived and designed the experiments; G.S carried out setaria transformation, plant cultivation and genotyping; J.L-F carried out all of the remaining experiments and analysed the data. J.L-F and J.A.L 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 S3D; Sophie Johnson, Chiara Perico, Daniela Vlad, Sovanna Tan, Thomas Hughes and Maricris Zaidem for discussion throughout the experimental work and during manuscript preparation.

COMPETING INTERESTS

No competing interests are declared.

FUNDING

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_4 Rice grant awarded to the University of Oxford (2015-2019; OPP1129902).

REFERENCES

Bravo-Garcia, A., Yasumura, Y., and Langdale, J.A., 2009. Specialization of the Golden2-like regulatory pathway during land plant evolution. *New Phytologist*, 183 (1), 133–141.

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., Chen, S.C.-C., Chen, S.M., Li, W.-H., and Ku, M.S.B., 2012. Characterizing Regulatory and Functional Differentiation between Maize Mesophyll and Bundle Sheath Cells by Transcriptomic Analysis. *Plant Physiology*, 160 (1), 165–177.

Chazaux, M., Schiphorst, C., Lazzari, G., and Caffarri, 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 (6), 1630–1648.

Chiang, Y.-H., Zubo, Y.O., Tapken, W., Kim, H.J., Lavanway, A.M., Howard, L., Pilon, M., Kieber, J.J., and Schaller, G.E., 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 (1), 332–348.

Christin, P.-A. and Osborne, C.P., 2013. The recurrent assembly of C4 photosynthesis, an evolutionary tale. *Photosynthesis research*, 117 (1–3), 163–75.

Concordet, J.-P. and Haeussler, M., 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, 46 (W1), W242–W245.

Cribb, L., Hall, L., and Langdale, J., 2001. Four Mutant Alleles Elucidate the Role of the G2 Protein in

562 the Development of C4 and C3 Photosynthesizing Maize Tissues. *Genetics*, 159, 787–797.
563 Van Eck, J. and Startwood, K., 2015. Setaria viridis. *In*: K. Wang, ed. *Agrobacterium Protocols*. Springer,
564 417.
565 Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S., 2009. Golden gate shuffling: a one-pot DNA
566 shuffling method based on type IIs restriction enzymes. *PloS one*, 4 (5), e5553.
567 Engler, C. and Marillonnet, S., 2013. Combinatorial DNA assembly using golden gate cloning. *Methods*
568 *in Molecular Biology*, 1073, 141–156.
569 Fitter, D.W., Martin, D.J., Copley, M.J., Scotland, R.W., and Langdale, J.A., 2002. GLK gene pairs
570 regulate chloroplast development in diverse plant species. *Plant Journal*, 31 (6), 713–727.
571 Gallaher, T.J., Peterson, P.M., Soreng, R.J., Zuloaga, F.O., Li, D.-Z., Clark, L.G., Tyrrell, C.D., Welker,
572 C.A.D., Kellogg, E.A., and Teisher, J.K., 2022. Grasses through space and time: An overview of the
573 biogeographical and macroevolutionary history of Poaceae. *Journal of Systematics and Evolution*,
574 60 (3), 522–569.
575 Hall, L.N., Rossini, L., Cribb, L., and Langdale, J.A., 1998. GOLDEN 2: A novel transcriptional regulator
576 of cellular differentiation in the maize leaf. *Plant Cell*, 10 (6), 925–936.
577 Hua, L., Stevenson, S.R., Reyna-Llorens, I., Xiong, H., Kopriva, S., and Hibberd, J.M., 2021. The bundle
578 sheath of rice is conditioned to play an active role in water transport as well as sulfur assimilation
579 and jasmonic acid synthesis. *Plant Journal*, 107 (1), 268–286.
580 Hudson, D., Guevara, D.R., Hand, A.J., Xu, Z., Hao, L., Chen, X., Zhu, T., Bi, Y.-M., and Rothstein, S.J.,
581 2013. Rice Cytokinin GATA Transcription Factor1 Regulates Chloroplast Development and Plant
582 Architecture. *Plant Physiology*, 162 (1), 132–144.
583 John, C.R., Smith-Unna, R.D., Woodfield, H., Covshoff, S., and Hibberd, J.M., 2014. Evolutionary
584 convergence of cell-specific gene expression in independent lineages of C4 grasses. *Plant*
585 *Physiology*, 165 (1), 62–75.
586 Kenrick, P. and Crane, P.R., 1997. The origin and early evolution of plants on land. *Nature*, 389 (6646),
587 33–39.
588 Kobayashi, K., Baba, S., Obayashi, T., Sato, M., Toyooka, K., Keränen, M., Aro, E.M., Fukaki, H., Ohta,
589 H., Sugimoto, K., and Masudaa, T., 2012. Regulation of root greening by light and auxin/cytokinin
590 signaling in Arabidopsis. *Plant Cell*, 24 (3), 1081–1095.
591 Lambret-Frotté, J., Almeida, L.C.S. De, Moura, S.M. De, and Souza, F.L.F., 2015. Validating Internal
592 Control Genes for the Accurate Normalization of qPCR Expression Analysis of the Novel Model
593 Plant Setaria viridis, 1–22.
594 Langdale, J.A., 2011. C4 cycles: past, present, and future research on C4 photosynthesis. *The Plant*
595 *Cell*, 23 (11), 3879–92.
596 Langdale, J.A. and Kidner, C.A., 1994. Bundle sheath defective, a mutation that disrupts cellular
597 differentiation in maize leaves. *Development*, 120 (3), 673–681.
598 Li, P. and Brutnell, T.P., 2011. Setaria viridis and Setaria italica, model genetic systems for the Panicoid

599 grasses. *Journal of Experimental Botany*, 62 (9), 3031–7.
600 Li, P., Ponnala, L., Gandotra, N., Wang, L., Si, Y., Tausta, S.L., Kebrom, T.H., Provart, N., Patel, R., Myers,
601 C.R., Reidel, E.J., Turgeon, R., Liu, P., Sun, Q., Nelson, T., and Brutnell, T.P., 2010. The
602 developmental dynamics of the maize leaf transcriptome. *Nature Genetics*, 42 (12), 1060–7.
603 Li, X., Wang, P., Li, J., Wei, S., Yan, Y., Yang, J., Zhao, M., Langdale, J.A., and Zhou, W., 2020. Maize
604 GOLDEN2-LIKE genes enhance biomass and grain yields in rice by improving photosynthesis and
605 reducing photoinhibition. *Communications Biology*, 3 (1), 151.
606 Lin, Y., Zhang, C., Lan, H., Gao, S., Liu, H., Liu, J., Cao, M., Pan, G., Rong, T., and Zhang, S., 2014.
607 Validation of potential reference genes for qPCR in maize across abiotic stresses, hormone
608 treatments, and tissue types. *PloS one*, 9 (5), e95445.
609 Marshall, D.M., Muhaidat, R., Brown, N.J., Liu, Z., Stanley, S., Griffiths, H., Sage, R.F., and Hibberd, J.M.,
610 2007. Cleome, a genus closely related to Arabidopsis, contains species spanning a developmental
611 progression from C3 to C4 photosynthesis. *Plant Journal*, 51 (5), 886–896.
612 Nguyen, C. V., Vrebalov, J.T., Gapper, N.E., Zheng, Y., Zhong, S., Fei, Z., and Giovannoni, J.J., 2014.
613 Tomato GOLDEN2-LIKE transcription factors reveal molecular gradients that function during fruit
614 development and ripening. *Plant Cell*, 26 (2), 585–601.
615 Nickelsen, J. and Rengstl, B., 2013. Photosystem II Assembly: From Cyanobacteria to Plants. *Annual*
616 *Review of Plant Biology*, 64 (1), 609–635.
617 Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic*
618 *Acids Research*, 29 (9), e45.
619 Porra, R.J., Thompson, W.A., and Kriedemann, P.E., 1989. Determination of accurate extinction
620 coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four
621 different solvents: verification of the concentration of chlorophyll standards by atomic absorption
622 spectroscopy. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 975 (3), 384–394.
623 Riechmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C., Keddie, J., Adam, L., Pineda, O., Ratcliffe,
624 O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman,
625 B.K., and Yu, G., 2000. Arabidopsis transcription factors: genome-wide comparative analysis
626 among eukaryotes. *Science (New York, N.Y.)*, 290 (5499), 2105–10.
627 Ritz, C. and Spiess, A.-N., 2008. qpcR: an R package for sigmoidal model selection in quantitative real-
628 time polymerase chain reaction analysis. *Bioinformatics*, 24 (13), 1549–1551.
629 Rossini, L., Cribb, L., Martin, D.J., and Langdale, J.A., 2001. The maize Golden2 gene defines a novel
630 class of transcriptional regulators in plants. *Plant Cell*, 13 (5), 1231–1244.
631 Safi, A., Medici, A., Szponarski, W., Ruffel, S., Lacombe, B., and Krouk, G., 2017. The world according
632 to GARP transcription factors. *Current Opinion in Plant Biology*, 39, 159–167.
633 Sagan, L., 1967. On the origin of mitosing cells. *Journal of Theoretical Biology*, 14 (3), 225-IN6.
634 Sage, R.F., Christin, P.-A., and Edwards, E.J., 2011. The C(4) plant lineages of planet Earth. *Journal of*
635 *Experimental Botany*, 62 (9), 3155–69.

636 Shapiro, S.S. and Wilk, M.B., 1965. An analysis of variance test for normality (complete samples).
637 *Biometrika*, 52 (3–4), 591–611.
638 Tausta, L.S., Li, P., Si, Y., Gandotra, N., Liu, P., Sun, Q., Brutnell, T.P., and Nelson, T., 2014.
639 Developmental dynamics of Kranz cell transcriptional specificity in maize leaf reveals early onset
640 of C4-related processes. *Journal of Experimental Botany*, 65 (13), 3543–3555.
641 Tu, X., Ren, S., Shen, W., Li, J., Li, Y., Li, C., Li, Y., Zong, Z., Xie, W., Grierson, D., Fei, Z., Giovannoni,
642 J., Li, P., and Zhong, S., 2022. Limited conservation in cross-species comparison of GLK
643 transcription factor binding suggested wide-spread cistrome divergence. *Nature*
644 *Communications*, 13 (1), 7632.
645 Tukey, J.W., 1949. Comparing Individual Means in the Analysis of Variance. *Biometrics*, 5 (2), 99–114.
646 Vlad, D., Abu-Jamous, B., Wang, P., and Langdale, J.A., 2019. A modular steroid-inducible gene
647 expression system for use in rice. *BMC Plant Biology*, 19 (1), 426.
648 Wang, P., Fouracre, J., Kelly, S., Karki, S., Gowik, U., Aubry, S., Shaw, M.K., Westhoff, P., Slamet-Loedin,
649 I.H., Quick, W.P., Hibberd, J.M., and Langdale, J.A., 2013. Evolution of GOLDEN2-LIKE gene
650 function in C3 and C4 plants. *Planta*, 237 (2), 481–495.
651 Washburn, J.D., Strable, J., Dickinson, P., Kothapalli, S.S., Brose, J.M., Covshoff, S., Conant, G.C.,
652 Hibberd, J.M., and Pires, J.C., 2021. Distinct C4 sub-types and C3 bundle sheath isolation in the
653 Paniceae grasses. *Plant Direct*, 5 (12), 1–14.
654 Waters, M.T. and Langdale, J.A., 2009. The making of a chloroplast. *The EMBO Journal*, 28 (19), 2861–
655 2873.
656 Waters, M.T., Moylan, E.C., and Langdale, J.A., 2008. GLK transcription factors regulate chloroplast
657 development in a cell-autonomous manner. *Plant Journal*, 56 (3), 432–444.
658 Waters, M.T., Wang, P., Korkaric, M., Capper, R.G., Saunders, N.J., and Langdale, J.A., 2009. GLK
659 Transcription Factors Coordinate Expression of the Photosynthetic Apparatus in Arabidopsis.
660 *Plant Cell*, 21 (4), 1109–1128.
661 Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S.,
662 Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., Ruhfel, B.R., Wafula, E., Der, J.P., Graham, S.W.,
663 Mathews, S., Melkonian, M., Soltis, D.E., Soltis, P.S., Miles, N.W., Rothfels, C.J., Pokorny, L., Shaw,
664 A.J., DeGironimo, L., Stevenson, D.W., Surek, B., Villarreal, J.C., Roure, B., Philippe, H.,
665 dePamphilis, C.W., Chen, T., Deyholos, M.K., Baucom, R.S., Kutchan, T.M., Augustin, M.M., Wang,
666 J., Zhang, Y., Tian, Z., Yan, Z., Wu, X., Sun, X., Wong, G.K.-S., and Leebens-Mack, J., 2014.
667 Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of*
668 *the National Academy of Sciences*, 111 (45), E4859–E4868.
669 Woo, K.C., Anderson, J.M., Boardman, N.K., Downton, W.J.S., Osmond, C.B., and Thorne, S.W., 1970.
670 Deficient Photosystem II in Agranal Bundle Sheath Chloroplasts of C4 Plants. *Proceedings of the*
671 *National Academy of Sciences*, 67 (1), 18–25.
672 Yasumura, Y., Moylan, E.C., and Langdale, J.A., 2005. A Conserved Transcription Factor Mediates

673 Nuclear Control of Organelle Biogenesis in Anciently Diverged Land Plants. *Plant Cell*, 17 (7),
674 1894–1907.
675 Zubo, Y.O., Blakley, I.C., Franco-Zorrilla, J.M., Yamburenko, M. V, Solano, R., Kieber, J.J., Loraine, A.E.,
676 and Schaller, G.E., 2018. Coordination of Chloroplast Development through the Action of the GNC
677 and GLK Transcription Factor Families. *Plant Physiology*, 178 (1), 130–147.
678
679

680 **FIGURE LEGENDS**

681 **Figure 1. Chloroplast development is not perturbed in *Zmglk1* mutants of maize**. **A)** Gene model
682 of *ZmGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated
683 mutagenesis in yellow and red below. The table shows the sequence of the wild-type (WT) and mutated
684 alleles with the predicted protein length. The gRNA sequence is highlighted in yellow and the PAM
685 sequence in red. **B-C)** Whole plant phenotype of wild-type (B) and a representative *Zmglk1* mutant 30
686 days after sowing (C). Scale bars correspond to 10 cm. **D-G)** Electron micrographs of chloroplast
687 ultrastructure in bundle sheath (BSC; D-E) and mesophyll (MC; F-G) cells of WT (D, F) and a
688 representative *Zmglk1* mutant (E, G). Scale bars correspond to 2 µm. **H-L)** Measurements of thylakoid
689 occupancy in mesophyll cells (H), thylakoid occupancy in bundle sheath cells (I), total chlorophyll (J),
690 PSII operating efficiency (K) and maximum quantum efficiency of PSII photochemistry (L). The WT
691 segregant is depicted in blue and the four independent *Zmglk1* lines are depicted in green. Each dot
692 represents a biological replicate, the square is the average value for that line and the bars are the
693 associated standard deviations. A minimum of 5 biological replicates were evaluated for thylakoid
694 occupancy and chlorophyll quantification, and 3 biological replicates were assessed for PSII
695 fluorescence. Letters indicate statistically different groups in comparison to WT; n.s. – no significant
696 statistical support for a difference from WT (pvalue≤0.05, one-way ANOVA and Tukey's HSD).

697

698 **Figure 2. Chloroplast development is not perturbed in *Svglk1* mutants of setaria. A)** Gene model
699 of *SvGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated
700 mutagenesis in yellow and red below. The table shows the sequence of the wild-type (WT) and mutated
701 alleles with the predicted protein length. The gRNA sequence is highlighted in yellow and the PAM
702 sequence in red. **B-D)** Whole plant phenotype of wild-type (B) and *Svglk1* mutants 20 days after sowing
703 (C-D). Scale bars correspond to 5 cm. **E-J)** Electron micrographs of chloroplast ultrastructure in bundle
704 sheath (BSC; E-G) and mesophyll (MC; H-J) cells of WT (E, H) and *Svglk1* mutants (F, G, I, J). Scale
705 bars correspond to 2 µm. **K-O)** Measurements of thylakoid occupancy in mesophyll cells (K), thylakoid
706 occupancy in bundle sheath cells (L), total chlorophyll (M), PSII operating efficiency (N) and maximum
707 quantum efficiency of PSII photochemistry (O). The WT segregant is depicted in blue and the two
708 independent *Svglk1* lines are depicted in green. Each dot represents a biological replicate, the square
709 is the average value for that line and the bars are the associated standard deviations. Twelve
710 chloroplasts were assessed for thylakoid occupancy quantification, and 5 biological replicates were
711 assessed for chlorophyll quantification and PSII fluorescence for each sample. n.s. - no significant
712 statistical support for a difference from WT (p-value≤0.05, one-way ANOVA and Tukey's HSD).

713

714 **Figure 3. Chloroplast development is perturbed in both mesophyll and bundle sheath cells of**
715 ***Svglk2* mutants of setaria. A)** Gene model of *SvGLK2* showing predicted exons in black boxes and
716 the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The table shows the

717 sequence of the wild-type (WT) and mutated alleles with the predicted protein length. The gRNA
718 sequence is highlighted in yellow and the PAM sequence in red. **B)** Whole plant phenotype of wild-
719 type, a heterozygous and a homozygous *Svglk2* mutant 20 days after sowing. Scale bar corresponds
720 to 5 cm. **C-D)** Confocal images of leaf cross sections showing overall chloroplast structure in WT (C)
721 and the *Svglk2* mutant (D). Bundle sheath cells are outlined in white. Scale bars correspond to 50 µm.
722 **E-H)** Electron micrographs of chloroplast ultrastructure in bundle sheath (BSC; E-F) and mesophyll
723 (MC; G-H) cells of WT (E, G) and the *Svglk2* mutant (F, H). Yellow arrows point to fragmented
724 membranes. Scale bars correspond to 2 µm. **I-K)** Measurements of thylakoid occupancy in bundle
725 sheath cells (I), thylakoid occupancy in mesophyll cells (J) and total chlorophyll (K). The WT segregant
726 is depicted in purple and the *Svglk2* line is depicted in pink. Each dot represents a biological replicate,
727 the square is the average value for that line and the bars are the associated standard deviations. Twelve
728 chloroplasts were assessed for thylakoid occupancy quantification, and 5 biological replicates were
729 assessed for chlorophyll quantification and PSII fluorescence for each sample. Letters indicate
730 statistically different groups in comparison to WT; n.s. - no significant statistical support for a difference
731 from WT (pvalue≤0.05, one-way ANOVA and Tukey's HSD).
732
733 **Figure 4. *GLK1* and *GLK2* transcript levels are reduced in *Svglk2* and *Zmg2* mutants of setaria**
734 **and maize. A-H)** Quantitative analysis of gene expression via qPCR in segregating wild-type (WT)
735 siblings and *glk* mutants of maize (A-D) and setaria (E-F). Relative normalized expression of *ZmGLK1*
736 (A), *ZmG2* (B), *ZmGNC* (C), *ZmCGA1* (D), *SvGLK1* (E) and *SvGLK2* (F) was calculated based on the
737 expression of two reference genes for each sample. Each dot represents a biological replicate, the
738 square is the mean value and the bars are the associated standard deviations. A minimum of 3 and 5
739 biological replicates were evaluated for maize and setaria samples, respectively. Statistical significance
740 was calculated using a Student's t-test and pairwise comparison with the respective WT sample. Letters
741 indicate statistically different groups in comparison to WT; n.s. - no significant statistical support for a
742 difference from WT (p-value≤0.05).
743
744 **Figure 5. *ZmGLK1* and *ZmG2* complement the *Svglk2* phenotype and activate transcription of**
745 **the setaria homologs**. **A-B)** Whole plant phenotype of *Svglk2* mutant overexpressing *ZmGLK1* (A) or
746 *ZmG2* (B) 30 days after sowing. Scale bars correspond to 5 cm. **C-F)** Quantitative analysis of gene
747 expression via qPCR of *ZmGLK1* (C), *ZmG2* (D), *SvGLK1* (E) and *SvGLK2* (F). Relative gene expression
748 was calculated based on the expression of two reference genes for each sample. Each dot represents
749 one of the 5 biological replicates evaluated, the square is the mean value and the bars are the
750 associated standard deviations. Statistical significance was calculated using a Student's t-test and
751 pairwise comparison with the respective wild-type (WT) sample. Letters indicate statistically different
752 groups in comparison to WT and *Svglk2* (p-value≤0.05).
753

754 **Figure 6. *ZmGLK1* and *ZmG2* are sufficient to induce dimorphic chloroplast development in both**
755 **mesophyll and bundle sheath cells of setaria**. **A-C)** Confocal images of leaf cross sections showing
756 overall chloroplast morphology in the *Svglk2* mutant (A, same as Figure 3D), and representative lines
757 expressing ZmGLK1 (B) or ZmG2 (C). Bundle sheath cells are outlined in white. Yellow arrows indicate
758 ectopic chloroplasts developing in vascular cells. Scale bars correspond to 50 µm. **D-I)** Electron
759 micrographs of chloroplast ultrastructure in bundle sheath (BSC; D-F) and mesophyll (MC; G-I) cells of
760 the *Svglk2* mutant (D-G; same as Figure 3G, H) and one representative line overexpressing *ZmGLK1*
761 (E-H) or ZmG2 (F-I). Scale bars correspond to 2 µm. **J-L)** Measurements of thylakoid occupancy in
762 mesophyll cells (J), thylakoid occupancy in bundle sheath cells (K) and total chlorophyll (L). The wild-
763 type (WT) segregant is depicted in purple, the *Svglk2* line in pink, the three independent lines
764 overexpressing *ZmGLK1* in orange and those overexpressing *ZmG2* in yellow. Each dot represents a
765 biological replicate, the square is the average value for that line and the bars are the associated
766 standard deviations. A minimum of 10 chloroplasts were evaluated for thylakoid occupancy
767 quantification and 5-10 biological replicates were used for chlorophyll quantification. Letters indicate
768 statistically different groups in comparison to WT and *Svglk2* (p-value$\leq$0.05, one-way ANOVA and
769 Tukey's HSD).

770 **SUPPLEMENTARY FIGURE LEGENDS**

771 **Figure S1. Phenotypic characterization of three independent *Zmglk1* mutant lines. A-C)** Whole
772 plant phenotype 30 days after sowing. Scale bars correspond to 10 cm. **D-F)** Transmission electron
773 microscopy (TEM) images showing mesophyll chloroplast ultrastructure. Scale bars correspond to 2
774 µm, 5 µm and 2 µm, respectively. **G-I)** TEM images showing bundle sheath chloroplast ultrastructure.
775 Scale bars correspond to 5 µm, 10 µm and 2 µm, respectively.

776

777 **Figure S2. Comparison of GLK sequences**. **A)** Phylogenetic reconstruction of the *GLK* gene family.
778 Protein sequences of the extensively characterised GLK members ZmGLK1, ZmG2, OsGLK1 and
779 OsGLK2 were used to create a HMMER profile that was used as a query to search for homologs in
780 whole protein databases of 11 species (*Zea mays, Setaria viridis, S. italica, Panicum virgatum, Sorghum*
781 *bicolor, Oryza sativa, Brachypodium distachyon, Vitis vinifera, Arabidopsis thaliana, Marchantia*
782 *polymorpha* and *Physcomitrium patens*). The presence of the MYB domain in those sequences was
783 confirmed using InterproScan, and only the sequences containing the conserved AREAEAA domain
784 characteristic of the GLK subfamily were selected. The resulting 25 amino acid sequences were aligned
785 using Clustal Omega. Phylogenetic reconstruction was performed using Bayesian inference on Mr.
786 Bayes using Poisson with fixed substitution rates as a model. Two independent MCMC runs with
787 1,000,000 generations were performed and no further improvement of log likelihood of prior
788 probabilities was observed. The tree was visualized on FigTree. Clade support on each node was
789 assessed using Bayesian posterior probabilities. Members of the GLK1 subgroup are depicted in green
790 and GLK2 in blue. **B)** Amino acid sequence alignment showing conservation between ZmGLK1,
791 SvGLK1, ZmG2 and SvGLK2. Magenta bars show the HLH associated with MYB or MYB-related
792 proteins; the AREAEAA conserved site is in blue and the last exon that contains the GCT-box domain
793 is in dark green.

794

795 **Figure S3. *GLK* transcript accumulation in bundle sheath and mesophyll cells. A-B)** Gene
796 expression in reads per kilobase million (RPKM) in mesophyll (MC) and bundle sheath cells (BSC) of
797 maize – *ZmGLK1* and *ZmG2* (A) and setaria – *SvGLK1* and *SvGLK2* (B). Maize expression data were
798 published by Chang et al. (2012) and setaria expression data were published by John et al. (2014). **C)**
799 Phylogenetic reconstruction based on 1000 bp upstream of the ATG in *GLK1* and *GLK2* genes.
800 Promoter sequences were aligned using MAFFT and phylogenetic reconstruction was performed using
801 MEGA software v. 7.0.26 and Maximum Likelihood based on the Tajima-Nei model. Clade support is
802 based on 1000 bootstrap replicates. **D)** Accumulation levels of *GLK1* and *GLK2* in transcripts per million
803 (TPM) in three different species of Paniceae grasses (*Urochloa fusca*, *Panicum hallii* and *Digitaria*
804 *californica*). All species carry out C_4 photosynthesis, but each uses a different decarboxylation pathway
805 in the bundle sheath cells.

806

807 **Figure S4. Phenotypic characterization of two independent lines of *Svglk2* complemented with**
808 ***ZmGLK1* or *ZmG2*. A-B)** Schematic of the constructs used to express *ZmGLK1* (A) or *ZmG2* (B) in the
809 *Svglk2* background. *HygR* depicts the hygromycin phosphotransferase gene and $OsACT_{pro}$ and
810 $ZmUBI_{pro}$ represent the constitutive rice actin and maize ubiquitin promoters, respectively. LB and RB
811 refers to left and right borders. **C-J)** Transmission electron microscopy images showing mesophyll
812 (MC) and bundle sheath (BSC) cell chloroplast ultrastructure in lines overexpressing *ZmGLK1* (C-F) or
813 *ZmG2* (G-J). Scale bar sizes are stated on the figures. **K-N)** Confocal images of leaf cross sections from
814 lines overexpressing *ZmGLK1* (K-L) or *ZmG2* (M-N) showing morphology of chloroplasts. Bundle
815 sheath cells are outlined in white. Scale bars correspond to 50 µm. Yellow arrows indicate ectopic
816 chloroplast formation in vascular cells.

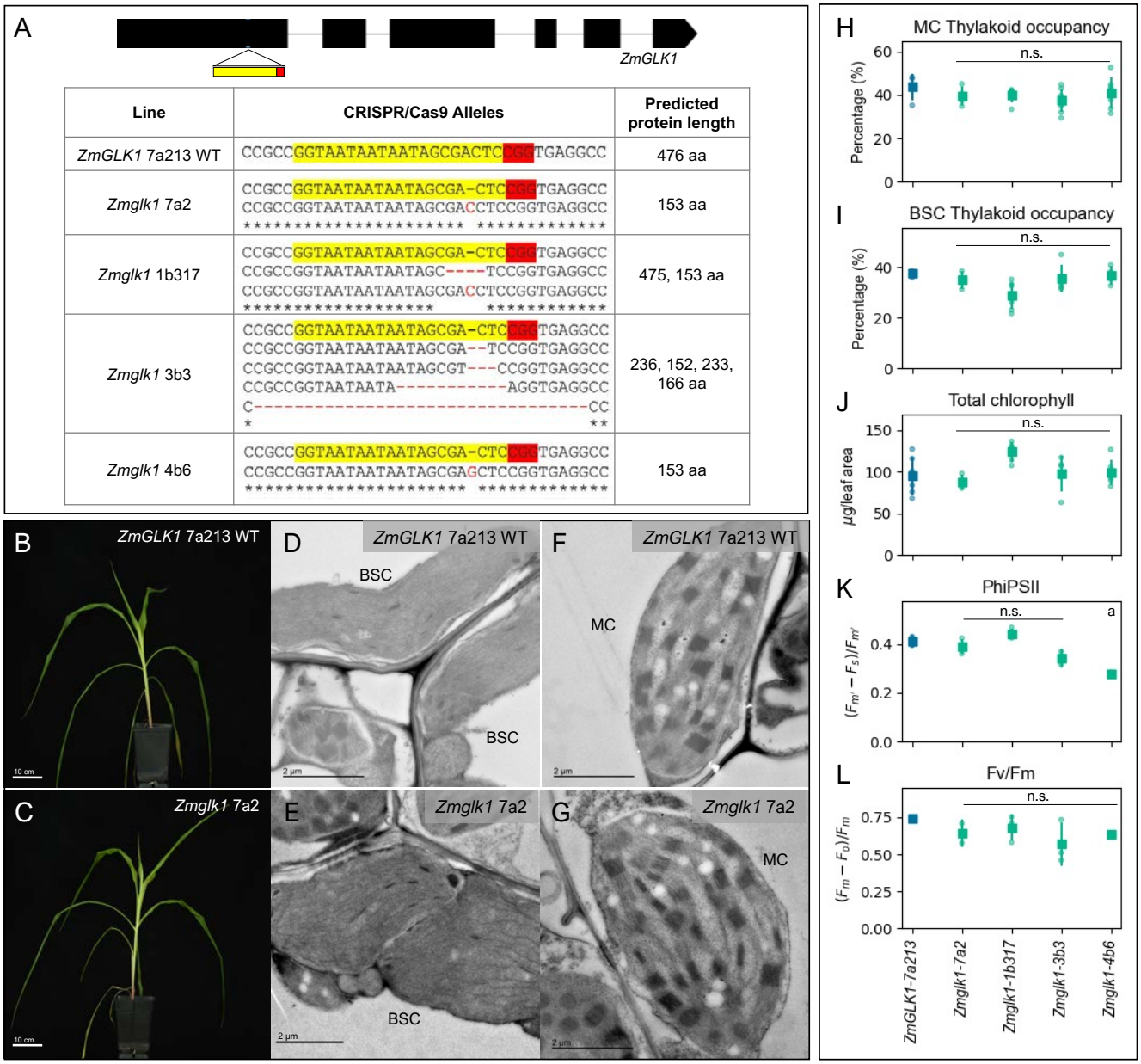
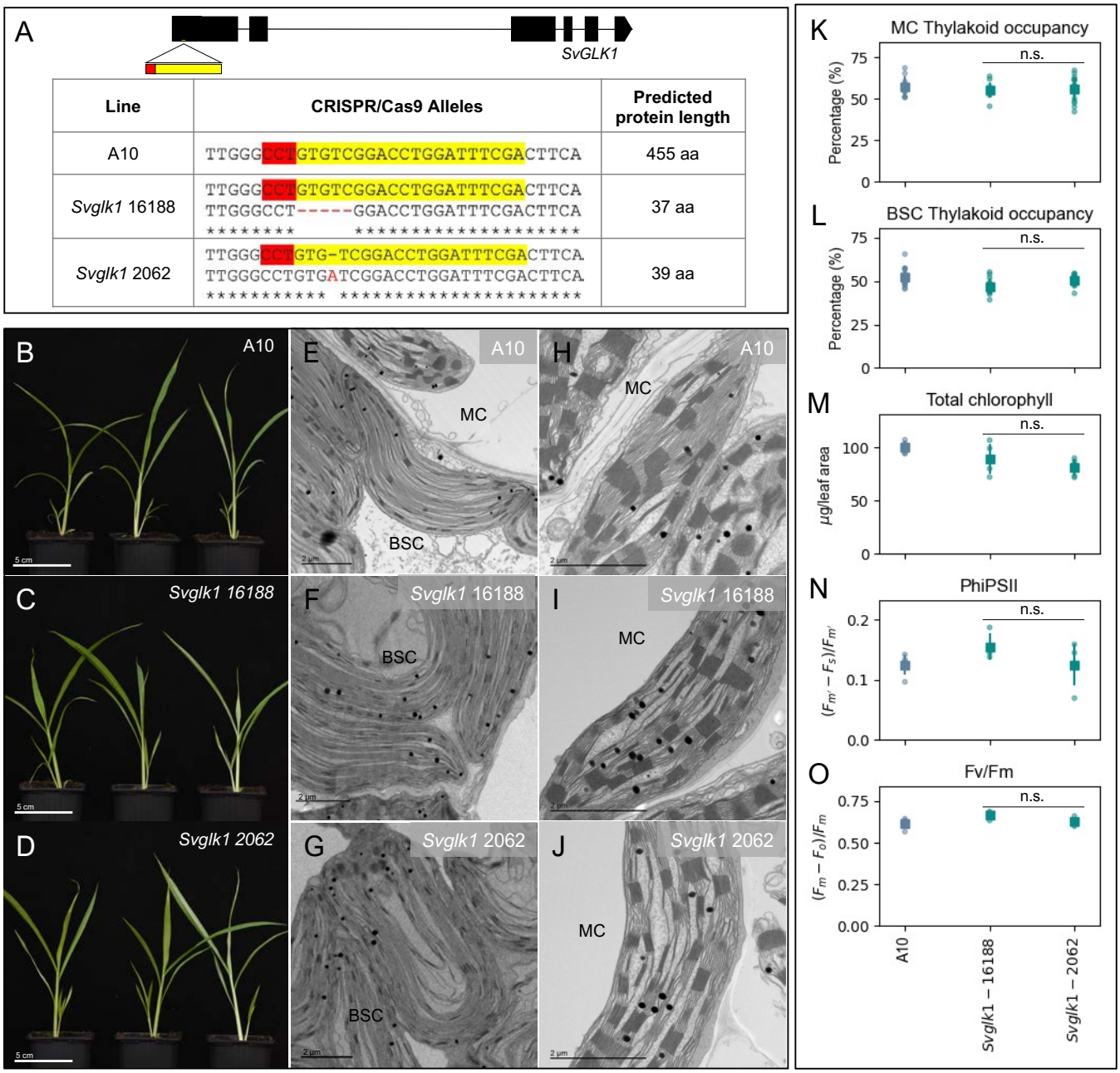


Figure 1. Chloroplast development is not perturbed in *Zmglk1* mutants of maize. **A)** Gene model of *ZmGLK1* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The table shows the sequence of the wild-type (WT) and mutated alleles 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 correspond to 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 correspond to 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. Letters indicate statistically different groups in comparison to WT; n.s. – no significant statistical support for a difference from WT (pvalue≤0.05, one-way ANOVA and Tukey's HSD).



Line	CRISPR/Cas9 Alleles	Predicted protein length
A10	TTGGGCCTGTGTCGGACCTGGATTTCGACTTCA	455 aa
Svglk1 16188	TTGGGCCTGTGTCGGACCTGGATTTCGACTTCA TTGGGCCT-----GGACCTGGATTTCGACTTCA ******** **********	37 aa
Svglk1 2062	TTGGGCCTGTG-TCGGACCTGGATTTCGACTTCA TTGGGCCTGTGATCGGACCTGGATTTCGACTTCA ********** **********	39 aa

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 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 correspond to 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 correspond to 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. n.s. - no significant statistical support for a difference from WT (p-value≤0.05, one-way ANOVA and Tukey's HSD).

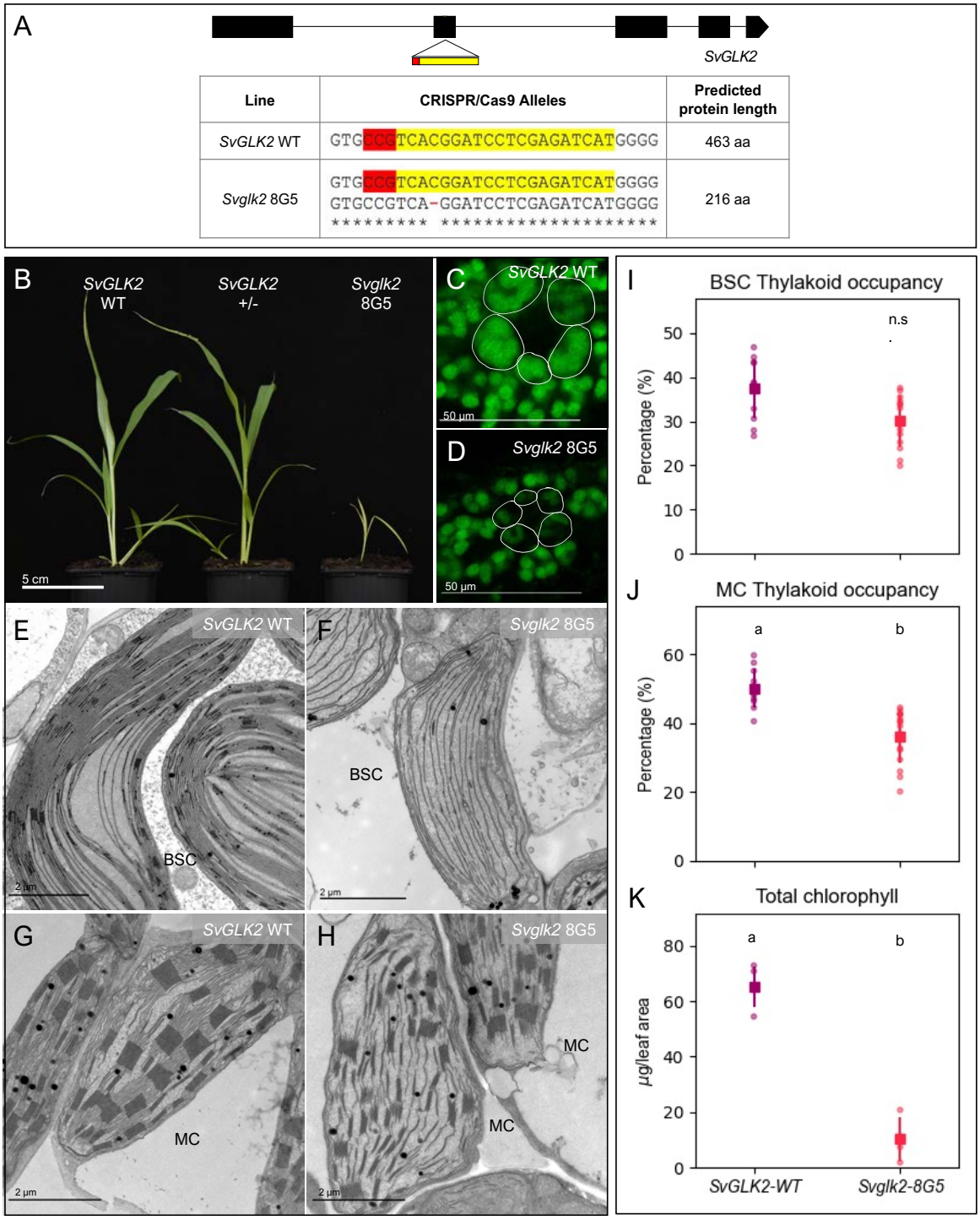


Figure 3. Chloroplast development is perturbed in both mesophyll and bundle sheath cells of *Svglk2* mutants of setaria. A) Gene model of *SvGLK2* showing predicted exons in black boxes and the target site of CRISPR/Cas9 mediated mutagenesis in yellow and red below. The table shows the sequence of the 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 corresponds to 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 correspond to 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 correspond to 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. Letters indicate statistically different groups in comparison to WT; n.s. - no significant statistical support for a difference from WT (pvalue≤0.05, one-way ANOVA and Tukey's HSD).

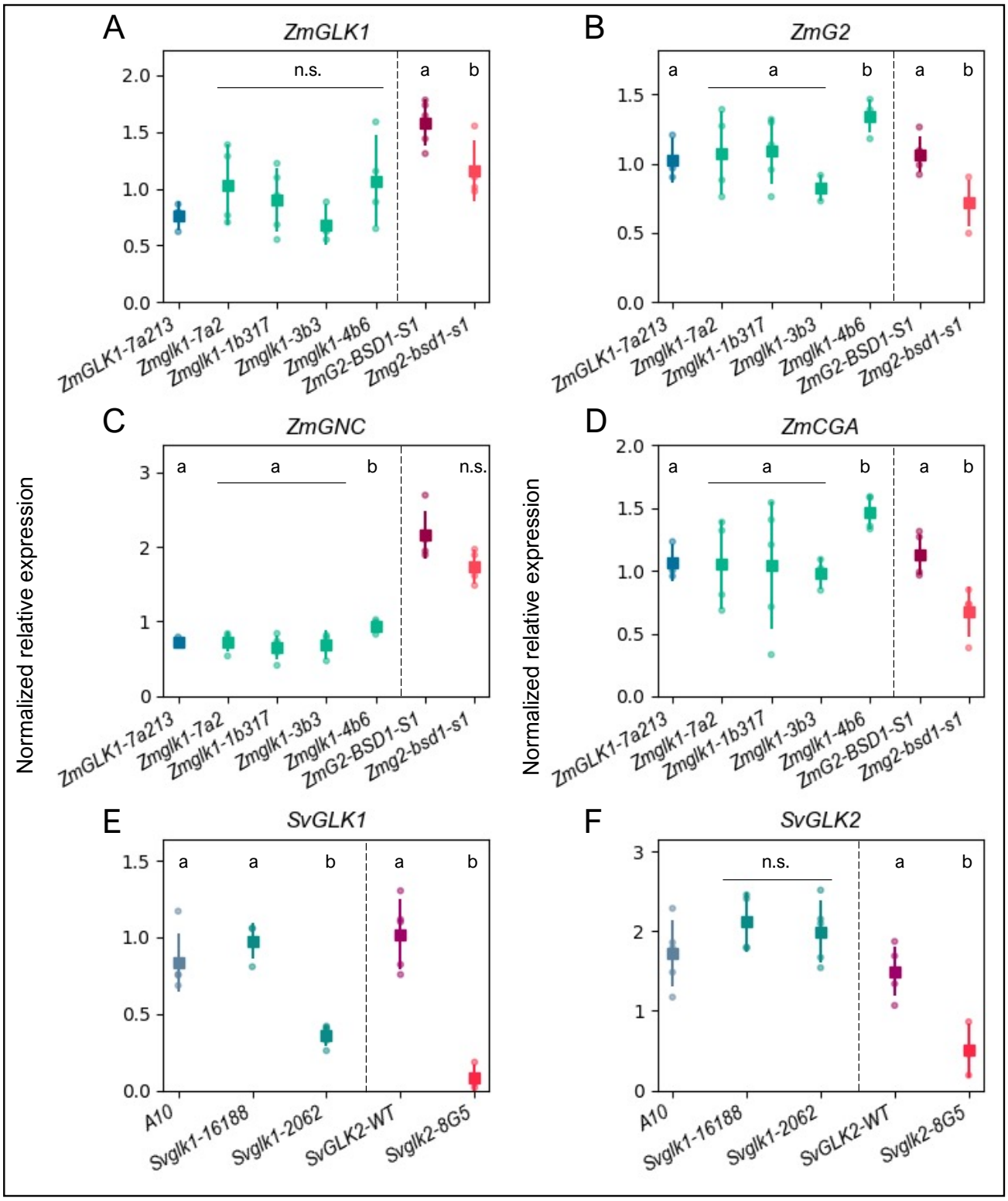


Figure 4. *GLK1* and *GLK2* transcript levels are reduced in *Svglk2* and *Zmg2* mutants of setaria and maize. A-H) Quantitative analysis of gene expression via qPCR in segregating wild-type (WT) siblings and *glk* mutants of maize (A-D) and setaria (E-F). Relative normalized expression of *ZmGLK1* (A), *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. Letters indicate statistically different groups in comparison to WT; n.s. - no significant statistical support for a difference from WT (p-value≤0.05).

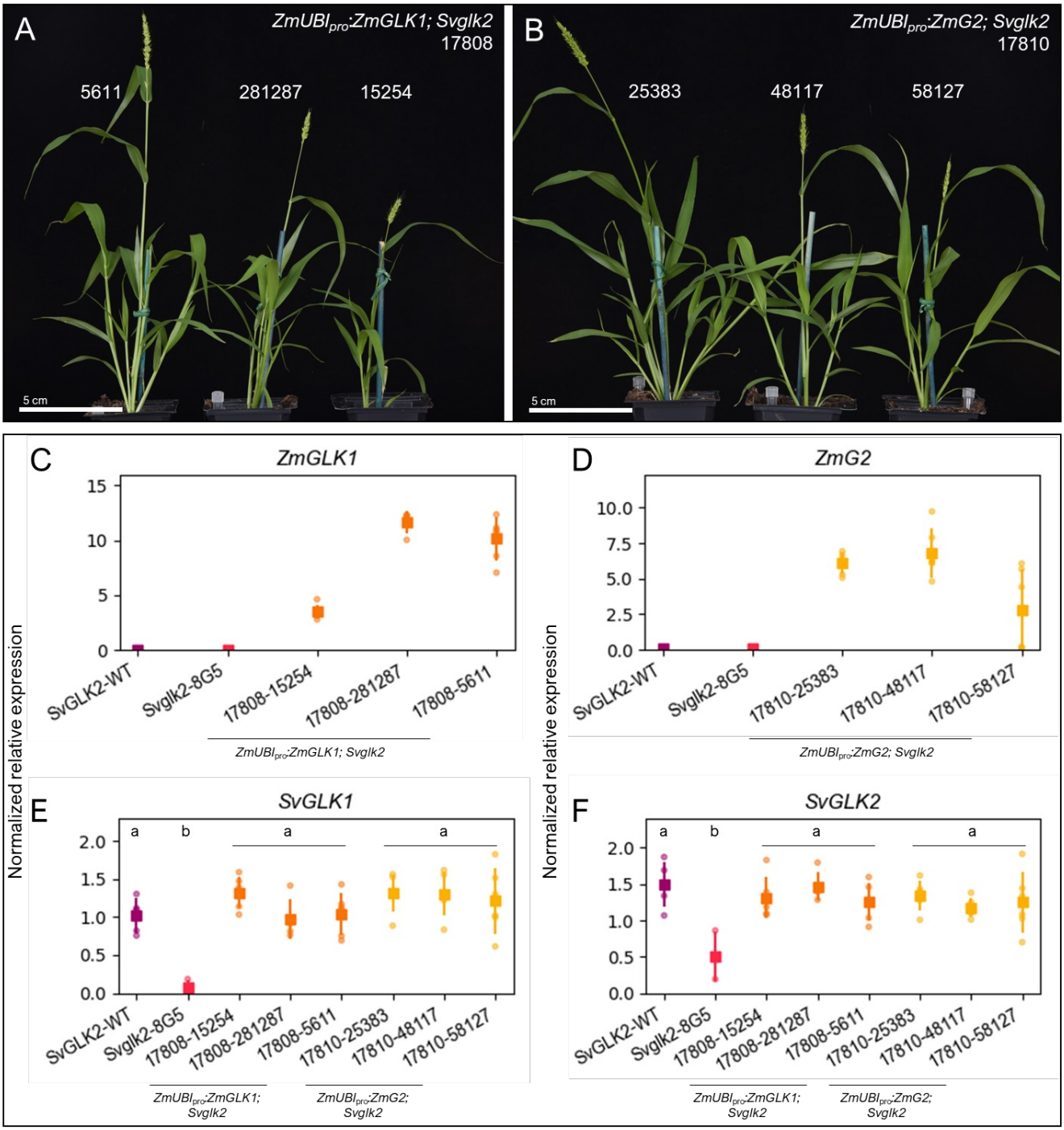


Figure 5. *ZmGLK1* and *ZmG2* complement the *Svglk2* phenotype and activate transcription of the setaria homologs. **A-B)** Whole plant phenotype of *Svglk2* mutant overexpressing *ZmGLK1* (A) or *ZmG2* (B) 30 days after sowing. Scale bars correspond to 5 cm. **C-F)** Quantitative analysis of gene expression 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. Letters indicate statistically different groups in comparison to WT and *Svglk2* (p-value≤0.05).

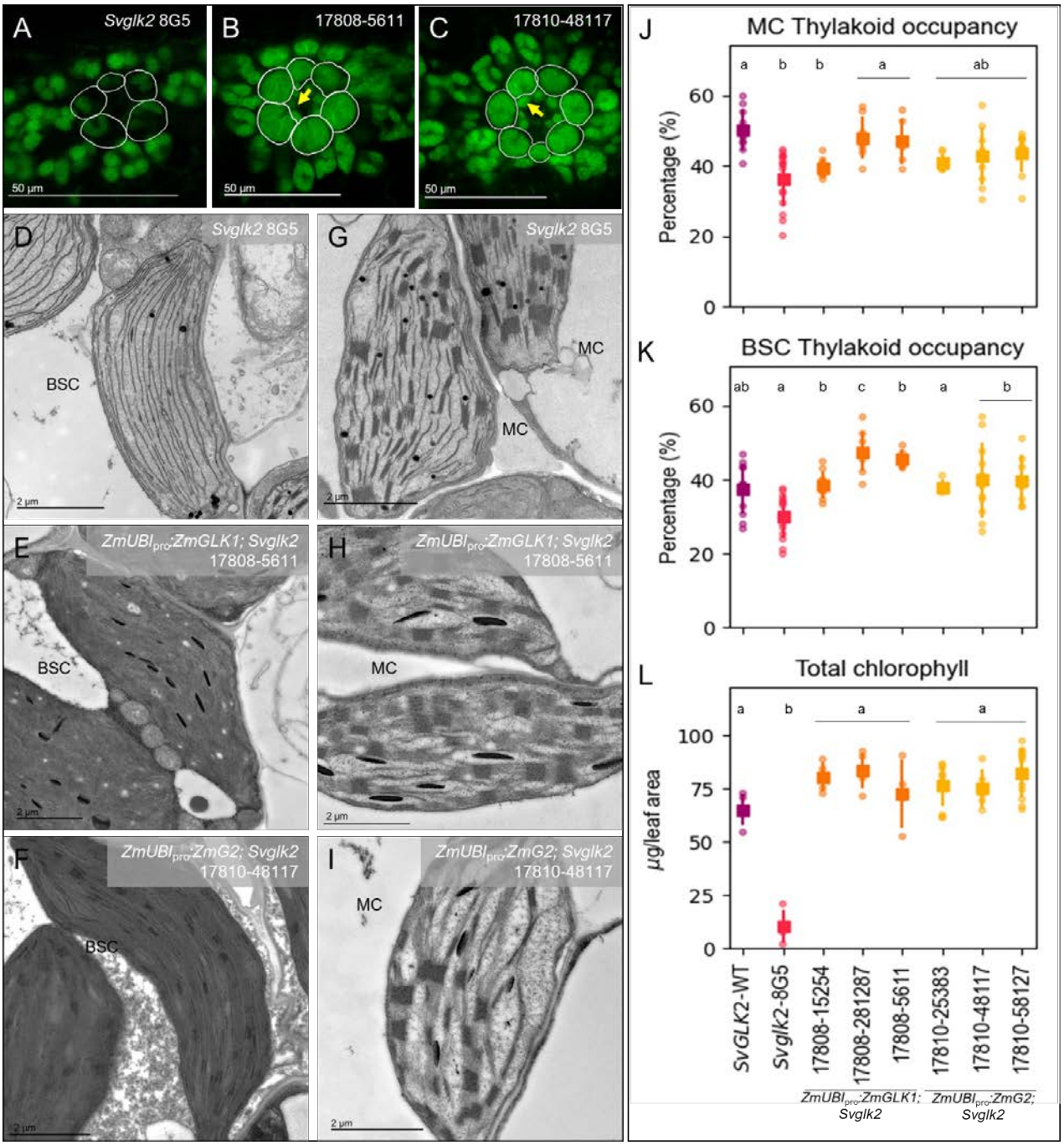


Figure 6. *ZmGLK1* and *ZmG2* are sufficient to induce dimorphic chloroplast development in both mesophyll and bundle sheath cells of setaria. A-C) Confocal images of leaf cross sections showing overall chloroplast morphology in the *Svglk2* mutant (A, same as Figure 3D), and representative lines expressing ZmGLK1 (B) or ZmG2 (C). Bundle sheath cells are outlined in white. Yellow arrows indicate ectopic chloroplasts developing in vascular cells. Scale bars correspond to 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 correspond to 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. Letters indicate statistically different groups in comparison to WT and *Svglk2* (p-value≤0.05, one-way ANOVA and Tukey's HSD).