

Identification and evolutionary analysis of the GOLDEN 2-LIKE gene family in foxtail millet

Huafeng chen

Shandong Agricultural University

Lei Qin

Shandong Agricultural University

Jinge Tian

National Maize Improvement Center, China Agricultural University

Xianglan Wang (✉ wangxianglan@sdau.edu.cn)

Shandong Agricultural University

Research Article

Keywords: GLK proteins, Phylogenetic analysis, Ortholog pairs, WGD, Cell type

Posted Date: July 8th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1814508/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

GOLDEN 2-LIKE (GLK) transcription factors are members of the *GARP* superfamily and are involved in chloroplast development and stress tolerance. These transcription factors have been studied at the genome level across several plant species. However, no study has comprehensively analyzed the *GLK* family in foxtail millet (*Setaria italica* L.). The present study discovered 59 *GLK* genes in foxtail millet. Multiple sequence alignment and gene motif analysis revealed the presence of Myb-SHAQKYF and Myb-CC-LHEQLE motifs in all the members. In addition, the gene promoters mainly had light- and stress-responsive cis-regulatory elements. Syntenic analysis showed that whole-genome duplications (WGD) and dispersed duplications contributed to the expansion of the *SiGLK* gene family. Gene expression analysis revealed that *SiGLKs* exhibited preferential expression in bundle sheath than in the mesophyll cells. Meanwhile, analyzing these genes in foxtail millet under abiotic stress demonstrated a significant role for *SiGLK30*, *38*, *48*, and *52* in regulating salt, cold, and drought stress responses. Finally, qRT-PCR confirmed that cold stress increased *SiGLK52* and *SiGLK31* expression. The study thus provides an overview of *GLK* gene evolution in foxtail millet and a clue to its function in bundle sheath and mesophyll cell and stress tolerance.

Introduction

C4 plants originated from C3 plants, and the process was reported to have occurred at least 66 times separately across several lineages (Shi et al. 2020; Sage et al. 2012). During the process, C4 plants have evolved a special leaf structure, known as the Kranz anatomy, consisting of two rings of cells around the vascular bundle of leaf veins in most C4 species, enabling them to maintain high carbon dioxide utilization to survive in low ambient CO₂ and high temperatures (Gowik et al. 2011; Osborne and Freckleton 2009; Edwards et al. 2001; Lundgren et al. 2014; Sedelnikova et al. 2018). Based on Kranz structure, ambient CO₂ was first fixed in OAA in the mesophyll (M) by phosphate pyruvate carboxylase (*PEPC*) and then transferred as malate into the bundle sheath (BS). With the help of *NADP-ME*, the fixed carbon dioxide was then liberated from malate and re-fixed by Rubisco. Meanwhile, enzymes involved in this process, such as carbonic anhydrase (*CA*), pyruvate orthophosphate dikinase (*PPDK*), phosphoenolpyruvate carboxykinase (*PCK*), malate dehydrogenase (*NADP-MDH*), and the previously mentioned phosphate pyruvate carboxylase (*PEPC*), *NADP*-dependent malic enzyme (*NADP-ME*), had a clear accumulation and enrichment.

GLK genes are a subset of the few transcription factors known to be involved in mesophyll cells (M) and bundle sheath (BS) compartmentalization in C4 plants (Rossini et al. 2001). *GLK* genes have two conserved domains, the Myb-DNA binding (containing the HLH motif) and the C-terminal domain (containing the conserved GCT box), as members of the *GARP* family (Riechmann et al. 2000).

Members of the *GLK* gene family have previously been found to regulate chloroplast formation and development conservatively. Mutations in the *golden2* gene resulted in the destruction of BS and pale green leaves in maize (Hall et al. 1998); in birch, deletion of *BpGLK1* reduced chlorophyll content and

disrupted chloroplast development (Gang et al. 2019); in tomato, suppression of *GLK1* resulted in pale leaves with reduced chlorophyll (Nguyen et al. 2014). In *P. patens*, *Arabidopsis* and rice, simultaneous loss of function of *GLK1* and *GLK2* resulted in pale leaves (Fitter et al. 2002; Wang et al. 2013; Bravo-Garcia et al. 2009; Yasumura et al. 2005). However, the two *GLK* genes regulated chloroplast development in different ways. In maize and sorghum, for example, loss of *GLK2* caused chloroplast disruption in BS cells but not in M cells. In non-C4 species, however, the two *GLK* genes acted redundantly to regulate chloroplast development (Wang et al. 2013; Yasumura et al. 2005; Bravo-Garcia et al. 2009; Chen et al. 2016). The fact that only double mutation of the two *GLKs* results in phenotypic difference from wild-type plants shows that *GLK1* and *GLK2* function separated after the C3-C4 split (Wang et al. 2013).

GLK genes have also been found to play a function in abiotic stress and disease resistance. In cotton, silencing *GhGLK1* caused more damage to plants under drought and cold stress than the wild type (Liu et al. 2021). Overexpression of *GLK1* in *Arabidopsis* conferred resistance to *Fusarium graminearum* (Savitch et al. 2007; Schreiber et al. 2011). Studies also indicated that *AtGLKs* exhibited tolerance to cucumber mosaic virus (Han et al. 2016). Furthermore, ectopic overexpression of *GLK1* conferred pathogen resistance in rice (Nakamura et al. 2009). *GLKs* were also involved in the control of leaf senescence (Qin et al. 2021).

In this study, we firstly explored *GLK* genes in the foxtail millet at the genome-wide level, and analyzed the subcellular localization, chromosomal distribution and gene structure. Of the 59 *GLK* genes identified in foxtail millet, *SiGLK30*, *38*, *48*, and *52* responded to drought, salt and cold treatments according to the stress analysis. This study will provide a reference for further research on the role of *GLKs* playing in foxtail millet.

Results

Identification of *GLK* genes in Foxtail millet

BLAST analysis using the *Arabidopsis* and maize *GLK* protein sequences (*GLK1*, *G2*) against the foxtail millet genome identified 59 *SiGLK* genes (Fig. 1, Table 1). The *SiGLK* proteins were 154 (*SiGLK53*) to 702 (*SiGLK29*) amino acids long, with an isoelectric point (*pI*) ranging from 5.03 (*SiGLK32*) to 10.02 (*SiGLK6*). The 59 *SiGLK* genes were found unevenly distributed across the nine chromosomes of foxtail millet, with genes ranging from 1 to 11 per chromosome; chromosome 4 had the most genes (11) and chromosome 8 had the least (1). These genes were renamed based on their position in the genome (Table 1). Subcellular localization analysis showed that all the *SiGLKs* are located in the nucleus (Table 1).

Table 1
List of the 59 SiGLK genes identified in *Setaria italica*.

Gene	Gene ID	Chr	pl	Mw	Protein	Subcellular Localization
SiGLK1	Seita.1G021500.1	chr1	5.94	56792.9	514	Nuclear
SiGLK2	Seita.1G061200.1	chr1	6.16	69271.5	635	Nuclear
SiGLK3	Seita.1G069400.1	chr1	8.68	39215.88	360	Nuclear
SiGLK4	Seita.1G092500.1	chr1	6.05	49130.81	450	Nuclear
SiGLK5	Seita.1G146600.1	chr1	8.63	43215.22	402	Nuclear
SiGLK6	Seita.1G268700.1	chr1	10.02	32426.15	298	Nuclear
SiGLK7	Seita.1G283000.1	chr1	8.85	38409.08	362	Nuclear
SiGLK8	Seita.1G286400.1	chr1	8.4	30178.24	276	Nuclear
SiGLK9	Seita.1G351300.1	chr1	6.24	73556.54	680	Nuclear
SiGLK10	Seita.2G013500.1	chr2	7.66	38181.93	359	Nuclear
SiGLK11	Seita.2G118400.1	chr2	5.08	45651.59	423	Nuclear
SiGLK12	Seita.2G168800.1	chr2	5.87	37990.49	354	Nuclear
SiGLK13	Seita.2G168900.1	chr2	6.02	34223.66	314	Nuclear
SiGLK14	Seita.2G197100.1	chr2	6.96	46776.73	465	Nuclear
SiGLK15	Seita.2G436900.1	chr2	6.79	38101.6	347	Nuclear
SiGLK16	Seita.3G029000.1	chr3	7.25	46810.79	434	Nuclear
SiGLK17	Seita.3G195600.1	chr3	7.82	24977.41	228	Nuclear
SiGLK18	Seita.3G199300.1	chr3	6.45	45850.3	413	Nuclear
SiGLK19	Seita.3G242400.1	chr3	8.57	40963.4	324	Nuclear
SiGLK20	Seita.3G370000.1	chr3	8.57	40963.4	391	Nuclear
SiGLK21	Seita.3G373000.1	chr3	7.02	39705.22	373	Nuclear
SiGLK22	Seita.4G050700.1	chr4	6.13	72711.52	667	Nuclear
SiGLK23	Seita.4G156200.1	chr4	5.14	49073.45	456	Nuclear
SiGLK24	Seita.4G175600.1	chr4	6.04	56924.86	537	Nuclear
SiGLK25	Seita.4G185800.1	chr4	7.86	36275.87	333	Nuclear
SiGLK26	Seita.4G198300.1	chr4	6.68	58981.22	541	Nuclear

Gene	Gene ID	Chr	pl	Mw	Protein	Subcellular Localization
SiGLK27	Seita.4G198500.1	chr4	6.06	74927.51	697	Nuclear
SiGLK28	Seita.4G198600.1	chr4	6.02	74956.4	697	Nuclear
SiGLK29	Seita.4G215900.1	chr4	5.46	75701.3	702	Nuclear
SiGLK30	Seita.4G266400.1	chr4	7.08	27077.27	243	Nuclear
SiGLK31	Seita.4G270600.1	chr4	5.61	51795.39	474	Nuclear
SiGLK32	Seita.4G273600.1	chr4	5.03	63914.46	586	Nuclear
SiGLK33	Seita.5G003100.1	chr5	5.98	49452.09	466	Nuclear
SiGLK34	Seita.5G024000.1	chr5	9.28	39834.07	378	Nuclear
SiGLK35	Seita.5G124700.1	chr5	5.47	51223.04	487	Nuclear
SiGLK36	Seita.5G379100.1	chr5	6.05	25015.69	240	Nuclear
SiGLK37	Seita.5G419200.1	chr5	5.12	64841.78	579	Nuclear
SiGLK38	Seita.5G468100.1	chr5	6.48	27132.39	254	Nuclear
SiGLK39	Seita.6G053400.1	chr6	6.87	37123.6	355	Nuclear
SiGLK40	Seita.6G120200.1	chr6	9.49	28309.26	263	Nuclear
SiGLK41	Seita.6G121500.1	chr6	5.95	33130.77	303	Nuclear
SiGLK42	Seita.6G152600.1	chr6	7.11	47034.86	456	Nuclear
SiGLK43	Seita.6G158900.1	chr6	5.58	30238.04	283	Nuclear
SiGLK44	Seita.7G055200.1	chr7	5.85	39487.8	346	Nuclear
SiGLK45	Seita.7G057700.1	chr7	5.46	40002.74	356	Nuclear
SiGLK46	Seita.7G163000.1	chr7	5.31	42251.23	391	Nuclear
SiGLK47	Seita.7G201600.1	chr7	9.54	32420.79	304	Nuclear
SiGLK48	Seita.7G223500.1	chr7	9.07	30743.22	279	Nuclear
SiGLK49	Seita.7G332300.1	chr7	5.77	41798.41	382	Nuclear
SiGLK50	Seita.8G019800.1	chr8	6.36	41091.22	379	Nuclear
SiGLK51	Seita.9G078300.1	chr9	7.73	38435.61	379	Nuclear
SiGLK52	Seita.9G080000.1	chr9	6.35	37380.33	350	Nuclear
SiGLK53	Seita.9G152500.1	chr9	8.84	16767.38	154	Nuclear

Gene	Gene ID	Chr	pl	Mw	Protein	Subcellular Localization
SiGLK54	Seita.9G165800.1	chr9	9.78	44161.68	408	Nuclear
SiGLK55	Seita.9G231000.1	chr9	5.76	38099.01	333	Nuclear
SiGLK56	Seita.9G416300.1	chr9	5.56	49138.76	449	Nuclear
SiGLK57	Seita.9G419000.1	chr9	6.79	38942.02	352	Nuclear
SiGLK58	Seita.9G485100.1	chr9	6.38	73746.87	688	Nuclear
SiGLK59	Seita.9G554300.1	chr9	8.94	41233.68	385	Nuclear

Further analysis based on the MEME predicted 15 potential motifs in the 59 *SiGLK* proteins (Fig. 2B, Table S1). The MEME annotation classified these proteins into six groups(I–VI) based on the composition of the motif (Fig. 2A). *GLK* gene-specific motifs 1 (Myb-SHLQKYF) and 2 (PELHRR) were found in I, II, III, IV, V, and VI groups, with 20, 15, 5, 6, 10, and 3 genes, respectively. All members of the group I contained one extra motif, the motif 3 (Myb-CC-LHEQLE) which distinguishes them from the other groups and is a symbol for the subfamily involved in phosphate starvation signaling (Kumar et al. 2017; Meng et al. 2021). Furthermore, group V was identified as the REC superfamily that acts as a phosphorylation-mediated switch within response regulators (RRs) with motifs 4, 5, 6, and 7. Meanwhile, group III contained an extra unannotated motif (motif 9).

Analysis of SiGLK Promoters

In silico analysis of the *SiGLK* promoter sequences (2000 bp upstream of the transcription start site, TSS) using plantcare discovered *cis*-acting elements associated with light and stress response as the most commonly distributed ones (Fig. 3). The light-responsive elements (LREs) identified included G-box, GT elements, I-box, H-box, and AT-rich sequences. The promoters of *SiGLK* genes had maximum G-box elements, ranging from 0 to 9. *SiGLK8* and *SiGLK17* contained the most G-boxes (9 and 6, respectively), indicating a significant influence of light on their expression. *SiGLK23*, *SiGLK41*, and *SiGLK43* lacked a G-box, while the other *SiGLKs* had at least one G-box.

Further analysis recognized ABRE and CGTCA as the most abundant defense and stress-responsive elements in the *SiGLK* promoters. *SiGLK8*, *SiGLK34*, and *SiGLK59* had the maximum ABREs (10, 9, and 7, respectively), and *SiGLK49*, *SiGLK43*, and *SiGLK12* contained the maximum number of CGTCA motifs (12, 12, and 12, respectively). These genes also had stress-responsive elements, such as LTR (low-temperature responsive) and MBS (drought-induced) elements. *SiGLK31*, *SiGLK52*, and *SiGLK40* contained the most LTRs, indicating their role or response under cold stress. In addition, almost all *SiGLKs* had one to three MBS, implying their potential to respond to drought.

Syntenic Analyses of SiGLKs in Foxtail millet

Collinearity analysis can provide information about gene duplication events and mode in plants. Therefore, we performed syntenic analysis to determine the gene duplication pairs and the method of duplication that led to *SiGLKs* during evolution. Five large-scale collinear blocks were discovered in the foxtail millet genome, including Chr1–Chr4, Chr1–Chr7, Chr2–Chr6, Chr2–Chr9, and Chr3–Chr5 (Fig. 4), consistent with previous studies in foxtail millet (Zhang et al. 2012). The analysis found that 27 *SiGLKs* were derived via WGD or segmental duplication and formed 15 syntenic duplication gene pairs (Table S2); 30 *SiGLKs* were derived via dispersed duplication, whereas two were derived from tandem duplication (Fig. 4). Among these, 25 *SiGLKs* were found within these large collinear blocks, and therefore, 14 of the 15 duplicated *SiGLK* pairs were considered to have expanded via WGD (Table S2).

Further, we calculated the *Ka/Ks* ratio (non-synonymous substitution rate, *Ka* and synonymous substitution rate, *Ks*) of the 15 duplicated *SiGLK* pairs using Tbttools (*Ka_Ks calculator*) to identify whether or not the *SiGLK* genes are under selective pressure. All duplicated *SiGLK* pairs except three pairs (*SiGLK15, 18*; *SiGLK15, 57*; *SiGLK18, 57*) underwent purifying selection because the *Ka_Ks* ratio ranged from 0.14 to 0.46 (Table S3). Meanwhile, the *Ks* of 5 of the 15 duplicated *SiGLK* pairs ranged from 0.6 to 0.8, and that of 6 of the 15 duplicated *SiGLK* pairs ranged from 1.0 to 1.7; only one pair had a *Ks* of 0.16.

Comparative Syntenic analysis Across Species

Further, BLAST analysis was carried out using the *GLK* protein sequences of Arabidopsis and maize (*GLK1, G2*) against the genomes of 26 species ranging from lower mosses to tall gymnosperms, from monocots to eudicots, and C3 to C4 plants (2 mosses, 1 lycophyte, 1 basal angiosperm, 2 gymnosperms, 11 monocots, and 9 eudicots) to investigate the expansion of *GLK* genes across the plants; the analysis discovered 1758 *GLKs* in total. Subsequently, a species tree was created using the 27 species with the NCBI taxonomy browser (Fig. 5). Excluding *Glycine max* (128) and

Musa acuminata (117), most species, from *Physcomitrella patens* to *Zea mays*, had similar numbers of *GLK* genes, demonstrating the importance of *GLK* in development regardless of mosses or angiosperms. However, the number of *GLK* genes appeared unrelated to genome size, particularly in gymnosperms like *Ginkgo biloba* and *Gnetum montanum*. *Ginkgo biloba*, with a 9555 Mb genome, has 38 *GLK* genes. The *Daucus carota* genome, on the other hand, of 422 Mb in size, contains 85 *GLK* genes. *Hordeum vulgare* had nearly the same number of *GLK* genes as maize, but it had a genome that was more than twice the size of maize. Furthermore, the closely related species, rice, sorghum, and maize, were chosen for comparing the synteny with foxtail millet using MCscanX (Fig. S1), considering Arabidopsis as the outgroup. We firstly discovered 71, 54, and 54 *GLK* genes in maize, sorghum, and rice, respectively, and performed syntenic analysis on each of them, as performed in foxtail millet (Fig. S2, Table S5). Except for a few uncertain *GLKs*, rice, maize, and sorghum have 30, 51, and 24 *GLKs*, respectively, originated from WGD or segmental duplication, and 19, 16, and 26 *GLKs* from dispersed duplication; only 3 *GLKs* (*SbGLK48, 49, 50*) derived from tandem duplication in sorghum, but none in rice or maize (Fig. S2D). Besides, 96, 77, 74, and 13 ortholog gene pairs were identified between foxtail millet and maize, rice, sorghum, and Arabidopsis, respectively (Table S5). Among these, 11 *SiGLKs* had at least one collinear gene in Arabidopsis, implying that they originated before the monocot-dicot split; 33 *SiGLKs* had at least

one collinear gene in rice, sorghum, and maize; and three *SiGLKs* had at least one collinear gene in sorghum and maize, but none in rice (Fig. S3); these observations imply that most *SiGLKs* originated before the C3-C4 split and after the dicot-monocot divergence.

Our analysis showed that each *GLK* in foxtail millet had two or more orthologs, probably due to gene duplication, resulting in additional paralogous genes. For example, *SiGLK15*, *SiGLK18*, and *SiGLK57* are paralogous and have two or three orthologs in rice. To further clarify the evolution of *GLK* genes, we investigated the genome collinearity of rice, foxtail millet, sorghum, and maize using Tbttools (Find Path by Gene Pair) (Chen et al. 2020) (Table S6). The analysis showed that most *SiGLK* genes corresponded to a particular orthologous gene in rice at the expected colinear location, such as *OsGLK30-SiGLK18-SbGLK43-ZmGLK47* and *OsGLK2-SiGLK33-SbGLK15-ZmGLK21* (Fig. 6). *SiGLK27* and *SiGLK28* were the only two *GLKs* produced by tandem duplication and occurred after foxtail millet-sorghum split, as both tandem genes shared the same rice and sorghum orthologs (*OsGLK36*, *SbGLK50*) at the expected colinear location. *SiGLK27* and *SiGLK28* were paralogous to *SiGLK2*, *SiGLK26*, and *SiGLK29* and formed a gene cluster (excluding *SiGLK2*). Further, the evolutionary history was determined by comparing them to the rice genome collinearity. According to our findings, one WGD event doubled *SiGLK26* and *SiGLK2*, then *SiGLK26* duplicated (dispersed duplication) to produce *SiGLK29*, and finally uneven crossing over produced *SiGLK27* and *SiGLK28* (Zhang 2003; Panchy et al. 2016).

Further, we calculated the time of divergence (mya) between these ortholog gene pairs between foxtail millet and rice, maize, sorghum based on synonymous substitution rate (*K_s*; Fig. S3) (Lynch and Conery 2000; Blanc and Wolfe 2004). The results indicated that the *K_s* within the syntenic gene pairs ranged from 0.2 to 0.4 (peak in 0.3), 0.2 to 0.6 (peak in 0.4), and 0.4 to 0.8 (peak in 0.6) in foxtail millet-sorghum, foxtail millet-maize, and foxtail millet-rice, respectively, consistent with the split of foxtail millet from sorghum (22.4 mya), maize (29.6 mya), and rice (47 mya), respectively (<http://www.timetree.org>) (Kumar et al. 2017).

Comparative Phylogenetic Analysis Across Species

We further aligned the amino acid sequences of 59 *SiGLKs* with rice, foxtail millet, sorghum, maize, and *Arabidopsis* to elucidate the evolutionary relationships between *GLK* genes (Fig. 7). A neighbor-joining tree constructed using the 279 *GLK* proteins with the MEGA software divided them into 11 groups (A–L), consistent with the grouping based on MEME analysis (Fig. S5). Group K (50) was the largest subfamily, with members containing motifs 1, 2, and 3. Group A was the second-largest, with members harboring motif 8 preceding motif 1 and motif 2; group A was followed by groups B (36), J (35), and L (29). Group H included all orthologs of maize *goden2* (*ZmGLK21*) and *GLK1* (*ZmGLK59*), *SiGLK33* and *SiGLK23* of foxtail millet, *OsGLK2* and *OsGLK33* of rice, *SbGLK15* and *SbGLK46* of sorghum, and *AtGLK13* and *AtGLK37* of *Arabidopsis thaliana*. *GLKs* from monocot plants (rice, foxtail millet, maize, and sorghum) clustered together rather than with *Arabidopsis* in each group.

Expression of SiGLKs in Different Foxtail millet Tissues and Cell types

We further investigated the transcriptome data (Zhang et al. 2012) obtained from four foxtail millet tissues (root, leaf, stem, and spica) to determine the expression patterns of the *SiGLK* genes (Fig. 8). Tbttools kit created an expression heatmap, and gene expression was normalized to determine the differences. The majority of *SiGLK* transcripts were detected in all four organs. *SiGLK52* had the highest expression in the root; *SiGLK23* had the highest expression in the leaf and was the only gene specifically expressed in leaves; *SiGLK17* and *SiGLK49* had the highest expression in the stem and spica, respectively. Meanwhile, *SiGLK1*, *SiGLK29*, *SiGLK36*, *SiGLK37*, *SiGLK46*, and *SiGLK56* were explicitly and highly expressed in roots. *SiGLK12*, *SiGLK15*, *SiGLK19*, *SiGLK21*, *SiGLK54*, and *SiGLK59* were found expressed preferentially in stems, whereas *SiGLK14*, *SiGLK42*, *SiGLK49*, *SiGLK50*, and *SiGLK53* were expressed preferentially in spica. Interestingly, we found that all highly expressed genes were members of groups I and III–V. Almost all group I members showed high expression in the four tissues, implying a broader role.

Foxtail millet is a C4 plant with a dimorphic chloroplast and photosynthesis compartmentalized into bundle sheath (BS) and mesophyll cells (M); therefore, the present study analyzed the expression of *SiGLKs* in M and BS separately (Fig. 9). Two sets of RNA-seq data from foxtail millet were analyzed (data1, PRJNA391551 (Dong et al. 2017); data2, PRJNA475365 (Cochrane et al. 2013)). The analysis of data1 revealed differential cellular accumulation for 39% of the *SiGLKs* (23 *SiGLKs*); 6 were *M*-enriched, and 17 were *BS*-enriched. In RNA-seq data2, 32% of the *SiGLKs* (19 *SiGLKs*) showed differential cellular accumulation; 3 were *M*-enriched, while 16 were *BS*-enriched. Meanwhile, two genes (*SiGLK23* and *SiGLK40*) showed *M* enrichment, and 12 showed *BS* enrichment in both datasets.

We also compared the cell-specific *GLKs* across four species. Rice, sorghum, and maize had 12, 26, and 17 *GLKs*, respectively, with differential cellular expression (Fig S6). The differentially expressed *BS*-enriched *GLK* genes in rice, sorghum, and maize were similar to those in foxtail millet. *SiGLK15*, *SiGLK18*, and *SiGLK57*, which are paralogous, were found expressed mainly in the *BS* as their rice (*OsGLK9*), sorghum (*SbGLK21*), and maize (*ZmGLK35*) orthologs. Similarly, the paralogs *SiGLK3* and *SiGLK25*, and their orthologs *OsGLK9*, *ZmGLK35*, and *SbGLK21* showed a *BS*-enriched expression pattern. *SiGLK33* demonstrated *BS*-biased expression consistent with its sorghum and maize orthologs (*SbGLK15* and *ZmGLK21*), but the opposite was true for its ortholog *OsGLK2* (*M*-biased) in rice.

Expression Pattern of SiGLKs Under Various Abiotic Stresses

The study further analyzed the expression patterns of *SiGLKs* during salt, drought, and cold stress using publicly available RNA-seq data (Yi et al. ; Meng et al. 2021) to elucidate the possible role of these genes in abiotic stress tolerance in foxtail millet (Fig. 10). Under drought stress, 24 *SiGLKs* showed significant differences in expression. Here, 12 of 24 *SiGLKs* were found upregulated, while the others were downregulated. *SiGLK19*, *SiGLK38*, *SiGLK48* and *SiGLK52* showed maximum upregulation in response to drought stress compared with the control samples, while *SiGLK18*, *SiGLK31* and *SiGLK47* showed maximum downregulation (Fig. 10A). Under salt stress, five *SiGLKs* were upregulated, whereas three were downregulated (Fig. 10B). *SiGLK52* showed maximum upregulation under salt stress. On the other hand,

SiGLK10 and *SiGLK31* were greatly reduced. Under cold stress, only *SiGLK23* displayed a rapid increase within 1 h (Fig. 10C). However, the number of cold-responsive *SiGLKs* gradually increased after 3 h. The genes *SiGLK23*, *SiGLK52*, *SiGLK3*, *SiGLK10*, and *SiGLK31* upregulated at 1, 3, 6, 16, and 24 h under cold stress. Meanwhile, at 3, 6, 16, and 24 h under cold stress, *SiGLK18*, *SiGLK27*, *SiGLK23*, and *SiGLK38* were downregulated (Fig. 10C). All three stress treatments resulted in the upregulation of four *SiGLKs* (*SiGLK30*, *38*, *48*, and *52*). Meanwhile, only *SiGLK18* was negatively regulated in response to all three stressors. *SiGLK52* had a higher expression than the other three upregulated *SiGLKs* under normal conditions and showed quick and continuous upregulation when exposed to stress (Fig. 10D). Further qRT-PCR confirmed the expression of *SiGLK52* and *31* during the cold treatment (Fig. 10D).

Discussion

Foxtail millet is one of the world's oldest cultivated crops and is widely grown in northern China as a major food crop (Barton et al. 2009). Moreover, as a typical C4 plant with a small genome (~ 500 Mb), foxtail millet has emerged as a convenient model plant for the study of C4 processes (Brutnell et al. 2010; Lata et al. 2013; Li and Brutnell 2011; Doust et al. 2009). *GLK* genes, governing plant cell differentiation in C4 plants and chloroplast formation in both C4 and C3 plants, have been discovered and described in maize, sorghum, rice, Arabidopsis, and various other plants (Wang et al. 2013; Fitter et al. 2002; Rossini et al. 2001; Liu et al. 2021; Nguyen et al. 2014). It can tolerate various abiotic stresses such as drought and low temperature because it prefers arid and semiarid environments (Bettinger et al. 2010). A thorough examination of the *GLK* gene family in foxtail millet will help us better understand this species' growth, development, and stress tolerance.

The present study discovered 59 *GLK* genes in foxtail millet and named them based on their chromosomal location. The number of *GLK* genes is similar to the C4 relatives, such as *Sorghum bicolor*, *Setaria viridis*, and *Panicum hallii*, but less than maize (*Zea mays*;71) due to one independent WGD that occurred in maize after the sorghum-maize split (Swigonová et al. 2004). Except for *Musa acuminata* and *Glycine max*, mosses to angiosperms had almost the same number of *GLKs*, highlighting the role of *GLK* in development in various members of the plant kingdom. We also found that the number of *GLKs* had little correlation with genome size. The gymnosperms *Gnetum montanum* and *Ginkgo biloba* with a large genome (9555 and 3653 Mb) contain fewer (38 and 27) *GLKs*, while *Brachypodium distachyon* has a tiny genome (271 Mb) with more (61) *GLKs*. Among the 27 species, foxtail millet has a moderately sized genome and *SiGLKs*. Detailed analysis revealed differences in the sequence length, relative molecular weight, and isoelectric point among the *SiGLK* proteins, which explained the *GLK* family's complexity. All *SiGLKs* were found localized in the nucleus, indicating their conserved transcriptional regulatory roles. Additionally, *SiGLKs* contained motifs 1 (Myb-SHLQKYF) and 2 (PELHRR), generally used to distinguish the *GLKs*. Meanwhile, motif3 (Myb-CC-LHEQLE), a type of MYB-like domain, was detected in all members of group I, the largest subfamily of *GLKs* in foxtail millet similar to maize (Liu et al. 2016).

Gene duplication has long been considered an important mechanism for acquiring numerous genes with various biological functions, and WGD operates as a key driving force allowing adaptive evolution to

proceed more efficiently (Wood et al. 2009; Soltis and Soltis 2016). Meanwhile, progress in genome sequencing techniques and the availability of various genomes have improved our understanding of plant evolution. Researchers have indicated that the foxtail millet ancestor lived through at least five WGD events (Fig. 4) (Wu et al. 2020; Liu et al. 2022; Zhang et al. 2020). However, the surge in *SiGLKs* mainly was attributed to the last WGD event in the foxtail millet, which occurred 70 mya (Zhang et al. 2012). Moreover, only 11 *SiGLK* genes had orthologous genes in *Arabidopsis thaliana*'s colinear region, and 33 *SiGLKs* had collinear genes in rice, which indicates that most *SiGLK* genes originated after the dicot-monocot divergence but before the foxtail millet-rice split. However, only one WGD (70 mya) was detected during the process, based on previously recognized WGDs (Zhang et al. 2020; Zhang et al. 2012).

The study also investigated the expression of *SiGLKs* in various tissues and two cell types (BS and M). *SiGLK52* had the highest expression in roots, although it was not explicitly expressed in this tissue. There are no reports on the role of *GLKs* in roots, so the importance of *SiGLK52* needs to be investigated; Meanwhile, *SiGLK23* and *SiGLK33* displayed the highest expression in the leaf. Furthermore, *SiGLK23* was only expressed in the leaf, whereas *SiGLK33* was expressed in the leaf, stem, and spica but not in the root. Spica showed a high and specific expression of *SiGLK49*. In addition, we discovered that 24% of *SiGLKs* showed differential accumulation between M and BS, implying the cell-specific accumulation as a general feature of the *GLK* gene family. When we analyzed various species, the same expression pattern was detected in maize (24%), sorghum (44%), and rice (20%). The *SiGLKs* and their orthologs in the C4 maize and sorghum had a similar cell-preferential expression but not in rice. This finding suggested a complex regulatory divergence in the evolution of C3-C4 plants. Moreover, the observations indicate the possible role of *SiGLK23* and *SiGLK33* in regulating cell differentiation and chloroplast formation, similar to the orthologous maize *GLK1* (*ZmGLK59*) and *goden2* (*ZmGLK59*) with similar expression patterns.

The study also discovered numerous *cis*-elements associated with light and abiotic stress, such as G-box, AREB, and LTR, in the promoters of *SiGLKs*. G-box was the most abundant light-related element. It was discovered as a light regulatory element in the small subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (*RBCS*) (Giuliano et al. 1988). Deletions or point mutations in G-box modulated the interaction with *GBF* (G-box binding factors) and influenced light-related gene expression (Donald and Cashmore 1990; Schindler et al. 1992). In addition, *SiGLK33* and *SiGLK23* are orthologs of maize *G2* and *GLK1* regulated by light; the study found more G-box elements in *SiGLK33* (7) than in *SiGLK23* (0), probably resulting in differences in their expression patterns. Meanwhile, all *SiGLKs*, except seven, had ABRE (ABA-responsive element), the primary *cis*-element for ABA-responsive gene expression (Nakashima and Yamaguchi-Shinozaki 2013), as the most common stress-related element, indicating the role of *SiGLKs* in stress responses. Further analysis of the gene expression under stress using publicly available transcriptome data showed that drought and cold treatment affected 24 and 23 *SiGLKs*, respectively, with 14 *SiGLKs* (5 upregulated and 9 downregulated) being differentially expressed under both drought and cold stress. The gene expression patterns were consistent with the widely distributed drought- and cold-responsive *cis*-elements, such as ABRE and MBS. Meanwhile, RNA-seq data obtained from samples under cold stress and qRT-PCR revealed that *SiGLK31* and *SiGLK52*, two of the three *SiGLKs* (*SiGLK* 31, 40, and

52) with the most LTRs, were upregulated. These findings confirmed the role of *SiGLKs* in stress response.

Conclusion

The present study identified and characterized 59 *GLK* genes in foxtail millet based on bioinformatic analysis. Syntenic analysis revealed WGD and dispersed duplication as the driving forces for the expansion of *SiGLK* genes. Analysis of the foxtail millet transcriptome revealed expression of these genes in almost all tissues, indicating their role in plant growth and development. Meanwhile, compartmentalized expression in the BS and M cells suggested their potential roles in dimorphic chloroplast development. Further, analysis of the expression of these genes in foxtail millet under abiotic stress demonstrated their roles in regulating the stress response. Among the various *SiGLKs*, *SiGLK30*, *38*, *48*, and *52* were implicated in salt, cold, and drought stress. Thus, the study's findings advance our understanding of *GLKs* and propose candidate genes for improving stress tolerance in foxtail millet. Further studies should focus on the functional analysis of *SiGLKs* for a comprehensive understanding of the stress regulatory pathways.

Materials And Methods

Identification of GLK Genes

The *GLK* protein sequences of Arabidopsis (<https://www.arabidopsis.org>) and maize (*GLK1*, *G2*) were used as query sequences to blast against the genomes of 26 species, including mosses (*Sphagnum magellanicum*, *Physcomitrium patens*), lycophytes (*Selaginella moellendorffii*), gymnosperms (*Gnetum montanum*, *Ginkgo biloba*), basal angiosperms (*Amborella trichopoda*), monocots (*Setaria italica*, *Zea mays*, *Oryza sativa*, *Sorghum bicolor*, *Acorus americanus*, *Musa acuminata*, *Ananas comosus*, *Panicum hallii*, *Setaria viridis*, *Brachypodium distachyon*, *Hordeum vulgare*) and eudicots (*Amaranthus hypochondriacus*, *Daucus carota*, *Vitis vinifera*, *Gossypium raimondii*, *Glycine max*, *Cicer arietinum*, *Manihot esculenta*, *Populus trichocarpa*, *Cucumis sativus*) (Goodstein et al. 2011). The datas were downloaded from the phytozome (<https://phytozome-next.jgi.doe.gov>) and ensemble (<https://plants.ensembl.org>) databases. As a result, sequences with a similarity of more than 30% and an E-value less than E^{-10} were retained. And the retained sequences were uploaded to the Swissport and Conserved Domain Database (<https://www.ncbi.nlm.nih.gov>) for further validation to ensure that the sequences belonged to the *GLK* family (Lu et al. 2020). *Setaria italica*, *Zea mays*, *Oryza sativa*, *Sorghum bicolor*, and *Arabidopsis* were chosen for multiple align, collinear comparison, and other data analysis, and the *GLK* genes of all 27 species were used to establish the species tree generated by searching in the NCBI taxonomy browser (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>). The genome sizes of the 27 species were calculated by Tbttools from the assembled genome files. The *pI* (isoelectric points) and *MW* (molecular weights) of the *GLK* proteins were calculated using the ExPASy website (<https://web.expasy.org>).

Multiple Sequence Alignment and Phylogenetic analysis

MEGAX11 (Kumar et al. 2018) aligned and phylogenetically assessed the amino acid sequences of *GLK* genes in Arabidopsis, maize, sorghum, foxtail millet, and rice. The neighbor-joining (NJ) tree was then built using 1000 bootstrap replicates. And the tree was annotated by iTOL (<https://itol.embl.de>).

Chromosomal Localization and Syntenic Analysis of GLK Genes in Foxtail millet, Maize, Sorghum and Rice

All discovered *GLK* genes were mapped to chromosomes using information received from the phytozome. MCscanX (Wang et al. 2012) was used for collinearity analysis. Tbtools (Chen et al. 2020) was used to visualize the collinearity relationship across foxtail millet, maize, rice, sorghum, and Arabidopsis thaliana. $T = Ks/2\lambda$ ($\lambda = 6.561029 \times 10^{-9}$) was used to calculate the divergence time (Lynch and Conery 2000; Blanc and Wolfe 2004).

Predicted Protein Motifs and Gene Structure Characterization GLK Genes in Foxtail millet

The protein motifs of *GLK* genes in foxtail millet were discovered and searched using the MEME (<https://meme-suite.org>). The results were visualized using Tbtools.

Cis-regulatory Element Analysis of SiGLKs in Foxtail millet

Plantcare (Lescot et al. 2002) was used to analyze the 2000pb promoter sequences (upstream of the coding region) of *SiGLKs* retrieved from the phytozome. After TATA-box and CAAT-box were deleted, Tbtools was used to visualize the *cis-acting* elements.

RNA-seq Data Processing

The NCBI (<https://www.ncbi.nlm.nih.gov>) has already published or made available all of the data used in this study. SRX128223 for root, SRX12825 for stem, SRX128224 for leaf, and SRX128226 for spica are the specific data for tissue expression analyses in foxtail millet (Zhang et al. 2012). PRJNA391551, PRJNA48621, PRJNA475365 for foxtail millet (Cochrane et al. 2013; Dong et al. 2017); PRJNA391551, PRJNA486213 for maize (Dong et al. 2017); PRJNA673407 for rice (Xiong et al. 2021); PRJEB11652 for sorghum (Emms et al. 2016) are the particular data for *M* and *BS*. PRJNA684771 (Yi et al.), PRJNA805389 (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA805389/>), PRJNA650146 (Meng et al. 2021) are the drought, salt, and cold data for foxtail millet.

RNA Extraction and qRT-PCR Analysis

The foxtail millet (*Setaria italica*) was grown in a greenhouse at a temperature of 28°C (16 hours of light/8 hours of darkness). Seedlings were given a cold treatment after three weeks. Leaves were collected at 0h, 6h, 12h, and 24h after being exposed to cold stress (4°C). Total RNA was extracted from

all of the samples in this study using TRIzol Reagent (Sigma, USA) according to the manufacturer's instructions. Two-Step RT-PCR SuperMix was used to conduct quantitative real-time (*qRT*) PCR.

Abbreviations

GLK, *GOLDEN 2-LIKE*; pI, isoelectric points; MW, molecular weight; NJ, neighbor-joining; BS, bundle sheath; M, mesophyll; WGD, whole genome duplication; mya, million years ago.

Declarations

Author Contributions HC, XW designed the study, HC performed the research, analyzed the data and wrote the paper. LQ, JT and XW modified the manuscript. All authors have read and approved the final manuscript.

Funding This work was supported by grants from the National Natural Science Foundation of China (31901552 and 32172013).

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Barton L, Newsome SD, Chen FH, Wang H, Guilderson TP, Bettinger RL (2009) Agricultural origins and the isotopic identity of domestication in northern China. *Proc Natl Acad Sci U S A* 106(14):5523–5528. doi:10.1073/pnas.0809960106
2. Bettinger RL, Barton L, Morgan C (2010) The origins of food production in north China: A different kind of agricultural revolution. *Evolutionary Anthropology: Issues News and Reviews* 19(1):9–21. doi:https://doi.org/10.1002/evan.20236
3. Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16(7):1667–1678. doi:10.1105/tpc.021345
4. Bravo-Garcia A, Yasumura Y, Langdale JA (2009) Specialization of the Golden2-like regulatory pathway during land plant evolution. *New Phytol* 183(1):133–141. doi:https://doi.org/10.1111/j.1469-8137.2009.02829.x
5. Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu XG, Kellogg E, Van Eck J (2010) *Setaria viridis*: a model for C4 photosynthesis. *Plant Cell* 22(8):2537–2544. doi:10.1105/tpc.110.075309
6. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* 13(8):1194–1202. doi:10.1016/j.molp.2020.06.009

7. Chen M, Ji M, Wen B, Liu L, Li S, Chen X, Gao D, Li L (2016) GOLDEN 2-LIKE Transcription Factors of Plants. *Front Plant Sci* 7. doi:10.3389/fpls.2016.01509
8. Cochrane G, Alako B, Amid C, Bower L, Cerdeño-Tárraga A, Cleland I, Gibson R, Goodgame N, Jang M, Kay S, Leinonen R, Lin X, Lopez R, McWilliam H, Oisel A, Pakseresht N, Pallreddy S, Park Y, Plaister S, Radhakrishnan R, Rivière S, Rossello M, Senf A, Silvester N, Smirnov D, Ten Hoopen P, Toribio A, Vaughan D, Zalunin V (2013) Facing growth in the European Nucleotide Archive. *Nucleic Acids Res* 41(Database issue):D30–35. doi:10.1093/nar/gks1175
9. Donald RG, Cashmore AR (1990) Mutation of either G box or I box sequences profoundly affects expression from the Arabidopsis rbcS-1A promoter. *Embo j* 9(6):1717–1726
10. Dong P, Tu X, Chu PY, Lü P, Zhu N, Grierson D, Du B, Li P, Zhong S (2017) 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Mol Plant* 10(12):1497–1509. doi:10.1016/j.molp.2017.11.005
11. Doust AN, Kellogg EA, Devos KM, Bennetzen JL (2009) Foxtail millet: a sequence-driven grass model system. *Plant Physiol* 149(1):137–141. doi:10.1104/pp.108.129627
12. Edwards GE, Franceschi VR, Ku MS, Voznesenskaya EV, Pyankov VI, Andreo CS (2001) Compartmentation of photosynthesis in cells and tissues of C(4) plants. *J Exp Bot* 52(356):577–590
13. Emms DM, Covshoff S, Hibberd JM, Kelly S (2016) Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C4 Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C4 Species. *Mol Biol Evol* 33(7):1796–1806. doi:10.1093/molbev/msw057
14. Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA (2002) GLK gene pairs regulate chloroplast development in diverse plant species. *Plant J* 31(6):713–727. doi:https://doi.org/10.1046/j.1365-313X.2002.01390.x
15. Gang H, Li R, Zhao Y, Liu G, Chen S, Jiang J (2019) Loss of GLK1 transcription factor function reveals new insights in chlorophyll biosynthesis and chloroplast development. *J Exp Bot* 70(12):3125–3138. doi:10.1093/jxb/erz128
16. Giuliano G, Pichersky E, Malik VS, Timko MP, Scolnik PA, Cashmore AR (1988) An evolutionarily conserved protein binding sequence upstream of a plant light-regulated gene. *Proc Natl Acad Sci U S A* 85(19):7089–7093. doi:10.1073/pnas.85.19.7089
17. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2011) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40(D1):D1178–D1186. doi:10.1093/nar/gkr944
18. Gowik U, Bräutigam A, Weber KL, Weber AP, Westhoff P (2011) Evolution of C4 photosynthesis in the genus *Flaveria*: how many and which genes does it take to make C4? *Plant Cell* 23(6):2087–2105. doi:10.1105/tpc.111.086264
19. Hall LN, Rossini L, Cribb L, Langdale JA (1998) GOLDEN 2: A Novel Transcriptional Regulator of Cellular Differentiation in the Maize Leaf. *Plant Cell* 10(6):925–936. doi:10.1105/tpc.10.6.925
20. Han X-Y, Li P-X, Zou L-J, Tan W-r, Zheng T, Zhang D-W, Lin H-H (2016) GOLDEN2-LIKE transcription factors coordinate the tolerance to Cucumber mosaic virus in Arabidopsis. *Biochem Biophys Res*

21. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* 35(6):1547–1549. doi:10.1093/molbev/msy096
22. Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* 34(7):1812–1819. doi:10.1093/molbev/msx116
23. Lata C, Gupta S, Prasad M (2013) Foxtail millet: a model crop for genetic and genomic studies in bioenergy grasses. *Crit Rev Biotechnol* 33(3):328–343. doi:10.3109/07388551.2012.716809
24. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res* 30(1):325–327. doi:10.1093/nar/30.1.325
25. Li P, Brutnell TP (2011) *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *J Exp Bot* 62(9):3031–3037. doi:10.1093/jxb/err096
26. Liu F, Xu Y, Han G, Zhou L, Ali A, Zhu S, Li X (2016) Molecular Evolution and Genetic Variation of G2-Like Transcription Factor Genes in Maize. *PLoS ONE* 11(8):e0161763–e0161763. doi:10.1371/journal.pone.0161763
27. Liu J, Mehari TG, Xu Y, Umer MJ, Hou Y, Wang Y, Peng R, Wang K, Cai X, Zhou Z, Liu F (2021) GhGLK1 a Key Candidate Gene From GARP Family Enhances Cold and Drought Stress Tolerance in Cotton. *Front Plant Sci* 12. doi:10.3389/fpls.2021.759312
28. Liu Y, Wang S, Li L, Yang T, Dong S, Wei T, Wu S, Liu Y, Gong Y, Feng X, Ma J, Chang G, Huang J, Yang Y, Wang H, Liu M, Xu Y, Liang H, Yu J, Cai Y, Zhang Z, Fan Y, Mu W, Sahu SK, Liu S, Lang X, Yang L, Li N, Habib S, Yang Y, Lindstrom AJ, Liang P, Goffinet B, Zaman S, Wegrzyn JL, Li D, Liu J, Cui J, Sonnenschein EC, Wang X, Ruan J, Xue JY, Shao ZQ, Song C, Fan G, Li Z, Zhang L, Liu J, Liu ZJ, Jiao Y, Wang XQ, Wu H, Wang E, Lisby M, Yang H, Wang J, Liu X, Xu X, Li N, Soltis PS, Van de Peer Y, Soltis DE, Gong X, Liu H, Zhang S (2022) The *Cycas* genome and the early evolution of seed plants. *Nat Plants* 8(4):389–401. doi:10.1038/s41477-022-01129-7
29. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 48(D1):D265–D268. doi:10.1093/nar/gkz991
30. Lundgren MR, Osborne CP, Christin P-A (2014) Deconstructing Kranz anatomy to understand C4 evolution. *J Exp Bot* 65(13):3357–3369. doi:10.1093/jxb/eru186
31. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–1155. doi:10.1126/science.290.5494.1151
32. Meng X, Liang Z, Dai X, Zhang Y, Mahboub S, Ngu DW, Roston RL, Schnable JC (2021) Predicting transcriptional responses to cold stress across plant species. *Proc Natl Acad Sci USA* 118(10):e2026330118. doi:10.1073/pnas.2026330118

33. Nakamura H, Muramatsu M, Hakata M, Ueno O, Nagamura Y, Hirochika H, Takano M, Ichikawa H (2009) Ectopic Overexpression of The Transcription Factor OsGLK1 Induces Chloroplast Development in Non-Green Rice Cells. *Plant Cell Physiol* 50(11):1933–1949. doi:10.1093/pcp/pcp138
34. Nakashima K, Yamaguchi-Shinozaki K (2013) ABA signaling in stress-response and seed development. *Plant Cell Rep* 32(7):959–970. doi:10.1007/s00299-013-1418-1
35. Nguyen CV, Vrebalov JT, Gapper NE, Zheng Y, Zhong S, Fei Z, Giovannoni JJ (2014) Tomato GOLDEN2-LIKE Transcription Factors Reveal Molecular Gradients That Function during Fruit Development and Ripening *The Plant Cell*. 26:585–601. 10.1105/tpc.113.118794. 2
36. Osborne CP, Freckleton RP (2009) Ecological selection pressures for C4 photosynthesis in the grasses. *Proc Biol Sci* 276(1663):1753–1760. doi:10.1098/rspb.2008.1762
37. Panchy N, Lehti-Shiu M, Shiu SH (2016) Evolution of Gene Duplication in Plants. *Plant Physiol* 171(4):2294–2316. doi:10.1104/pp.16.00523
38. Qin M, Zhang B, Gu G, Yuan J, Yang X, Yang J, Xie X (2021) Genome-Wide Analysis of the G2-Like Transcription Factor Genes and Their Expression in Different Senescence Stages of Tobacco (*Nicotiana tabacum* L.). *Front Genet* 12. doi:10.3389/fgene.2021.626352
39. Riechmann JL, Heard J, Martin G, Reuber L, Jiang C-Z, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G-L (2000) Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science* 290(5499):2105–2110. doi:doi:10.1126/science.290.5499.2105
40. Rossini L, Cribb L, Martin DJ, Langdale JA (2001) The maize golden2 gene defines a novel class of transcriptional regulators in plants. *Plant Cell* 13(5):1231–1244. doi:10.1105/tpc.13.5.1231
41. Sage RF, Sage TL, Kocacinar F (2012) Photorespiration and the evolution of C4 photosynthesis. *Annu Rev Plant Biol* 63:19–47. doi:10.1146/annurev-arplant-042811-105511
42. Savitch LV, Subramaniam R, Allard GC, Singh J (2007) The GLK1 ‘regulon’ encodes disease defense related proteins and confers resistance to *Fusarium graminearum* in Arabidopsis. *Biochem Biophys Res Commun* 359(2):234–238. doi:https://doi.org/10.1016/j.bbrc.2007.05.084
43. Schindler U, Menkens AE, Beckmann H, Ecker JR, Cashmore AR (1992) Heterodimerization between light-regulated and ubiquitously expressed Arabidopsis GBF bZIP proteins. *Embo j* 11(4):1261–1273
44. Schreiber KJ, Nasmith CG, Allard G, Singh J, Subramaniam R, Desveaux D (2011) Found in Translation: High-Throughput Chemical Screening in Arabidopsis thaliana Identifies Small Molecules That Reduce Fusarium Head Blight Disease in Wheat. *Mol Plant-Microbe Interactions®* 24(6):640–648. doi:10.1094/mpmi-09-10-0210
45. Sedelnikova OV, Hughes TE, Langdale JA (2018) Understanding the Genetic Basis of C(4) Kranz Anatomy with a View to Engineering C(3) Crops. *Annu Rev Genet* 52:249–270. doi:10.1146/annurev-genet-120417-031217
46. Shi W, Yue L, Guo J, Wang J, Yuan X, Dong S, Guo J, Guo P (2020) Identification and evolution of C(4) photosynthetic pathway genes in plants. *BMC Plant Biol* 20(1):132. doi:10.1186/s12870-020-

47. Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol* 30:159–165. doi:<https://doi.org/10.1016/j.pbi.2016.03.015>
48. Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. *Genome Res* 14(10a):1916–1923. doi:10.1101/gr.2332504
49. Wang P, Fouracre J, Kelly S, Karki S, Gowik U, Aubry S, Shaw MK, Westhoff P, Slamet-Loedin IH, Quick WP, Hibberd JM, Langdale JA (2013) Evolution of GOLDEN2-LIKE gene function in C3 and C4 plants. *Planta* 237(2):481–495. doi:10.1007/s00425-012-1754-3
50. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 40(7):e49–e49. doi:10.1093/nar/gkr1293
51. Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH (2009) The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* 106(33):13875–13879. doi:10.1073/pnas.0811575106
52. Wu S, Han B, Jiao Y (2020) Genetic Contribution of Paleopolyploidy to Adaptive Evolution in Angiosperms. *Mol Plant* 13(1):59–71. doi:10.1016/j.molp.2019.10.012
53. Xiong H, Hua L, Reyna-Llorens I, Shi Y, Chen K-M, Smirnov N, Kromdijk J, Hibberd JM (2021) Photosynthesis-independent production of reactive oxygen species in the rice bundle sheath during high light is mediated by NADPH oxidase. *Proc Natl Acad Sci USA* 118(25):e2022702118. doi:10.1073/pnas.2022702118
54. Yasumura Y, Moylan EC, Langdale JA (2005) A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell* 17(7):1894–1907. doi:10.1105/tpc.105.033191
55. Yi F, Huo M, Li J, Yu J Time-series transcriptomics reveals a drought-responsive temporal network and crosstalk between drought stress and the circadian clock in foxtail millet. *The Plant Journal* n/a(n/a). doi: <https://doi.org/10.1111/tpj.15725>
56. Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, Tao Y, Bian C, Han C, Xia Q, Peng X, Cao R, Yang X, Zhan D, Hu J, Zhang Y, Li H, Li H, Li N, Wang J, Wang C, Wang R, Guo T, Cai Y, Liu C, Xiang H, Shi Q, Huang P, Chen Q, Li Y, Wang J, Zhao Z, Wang J (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol* 30(6):549–554. doi:10.1038/nbt.2195
57. Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18(6):292–298. doi:[https://doi.org/10.1016/S0169-5347\(03\)00033-8](https://doi.org/10.1016/S0169-5347(03)00033-8)
58. Zhang X, Li X, Zhao R, Zhou Y, Jiao Y (2020) Evolutionary strategies drive a balance of the interacting gene products for the CBL and CIPK gene families. *New Phytol* 226(5):1506–1516. doi:10.1111/nph.16445

Figures

Figure 1

Location of 59 *GLK* genes in *Setaria italica*

Figure 2

Analysis of conserved motif and exon-intron structure of *GLK* genes in *Setaria italica*. (A) Phylogenetic analysis of *SiGLK* genes, (B) Analysis of conserved motif of *SiGLK* protein sequences, (C) Intron and exon analysis of the *SiGLK* gene.

Figure 3

Analysis of cis-regulatory elements in promoters of *SiGLK* genes.

Figure 4

Syntenic analysis in *Setaria italica*. The outer ring is the chromosome, the middle ring is the gene density matching the chromosome, and the inner ring is the syntenic block. Red lines highlight syntenic *GLK* gene pairs. *SiGLKs* are marked in black, red, blue and represent genes from dispersed duplication, WGD or segmental duplication and tandem. The legend indicates the number of genes per 100Kb.

Figure 5

Number of *GLK* genes and genome size of 27 species. Green bars indicate the number of *GLK* genes; red bars indicate the genome size (Mb) of each species.

Figure 6

Visualization of examples of corresponding *GLK* orthologs in maize, sorghum, millet, and rice.

Figure 7

Phylogenetic analysis of *GLK* proteins from Arabidopsis, maize, sorghum, foxtail millet and rice. Colored strips represent classification(A-L).

Figure 8

Heat-map of the expression pattern of *SiGLK* genes in four tissues namely leaf, root, stem and spica. The size of red circles represents the relative expression of *SiGLKs* (TPM).

Figure 9

Cellular accumulation expression analysis of *GLKs* in bundle sheath (BS) and mesophyll (M) cells in *Setaria italica*. DEGs (differential expression genes) between BS and M cells ($q < 0.05$, $|\log_2 \text{foldchange}| > 1$) are presented. A and C are from data1 (PRJNA391551); C and D are from data2 (PRJNA475365).

Figure 10

Stress responsiveness of *SiGLKs* in *Setaria italica*. (A) heatmap of *SiGLKs* using the $\log_2$ (fold-change) values of genes over all time points, CD represents drought treatment starting at the day compared with the control samples, (B) Differential expressed *SiGLKs* after salt treatment, (C) Up and down-regulated *SiGLKs* during a time-series of cold stress. The red rectangle represents upregulated *SiGLKs*, the blue rectangle represents downregulated *SiGLKs*, and the white rectangle indicates no-change *SiGLKs*, (D) qRT of *SiGLK52* and *SiGLK31* under cold stress.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryFiguresandTables.zip](#)