Identification of the Golden-2-like transcription factors gene family in *Gossypium hirsutum*

Zilin Zhao [Equal first author, 1, 2], **Jiaran Shuang** [Equal first author, 2], **Zhaoguo Li** [2], **Huimin Xiao** [2], **Yuling Liu** [2], **Tao Wang** [2], **Yangyang Wei** [2], **Shoulin Hu** [1], **Sumei Wan** [Corresp., 1], **Renhai Peng** [Corresp. 1, 2]

[1] College of Plant Science, Tarim University, Alar, Xinjiang, China

[2] Anyang Institute of Technology, Anyang, Henan, China

Corresponding Authors: Sumei Wan, Renhai Peng
Email address: wansumei510@163.com, aydxprh@163.com

Background. Golden2-Like (GLK) transcription factors are a type of transcriptional regulator in plants. They play a pivotal role in the plant physiological activity process and abiotic stress response.

Methods. In this study, the potential function of *GLK* family genes in *Gossypium hirsutum* was studied based on genomic identification, phylogenetic analysis, chromosome mapping and cis-regulatory element prediction. Gene expression of nine key genes were analyzed by qRT-PCR experiments.

Results. Herein, we identified a total of 146 *GhGLK* genes in *Gossypium hirsutum*, which were unevenly distributed on each of the chromosomes. There were significant differences in the number and location of genes between the At sub-genome and the Dt sub-genome. According to the phylogenetic analysis, they were divided into ten subgroups, each of which had very similar number and structure of exons and introns. Some cis-regulatory elements were identified through promoter analysis, including five types of elements related to abiotic stress response, five types of elements related to phytohormone and five types of elements involved in growth and development. Based on public transcriptome data analysis, we identified nine key *GhGLKs* involved in salt, cold, and drought stress. The qRT-PCR results showed that these genes had different expression patterns under these stress condition, suggesting that *GhGLK* genes played an important role in abiotic stress response. This study laid a theoretical foundation for the screening and functional verification of genes related to stress resistance of *GLK* gene family in cotton.

1 **Identification of the Golden-2-like transcription factors gene family in**

2 *Gossypium hirsutum*

3 Zilin Zhao[1,2†], Jiaran Shuang[2†], Zhaoguo Li[2], Huimin Xiao[2], Yuling Liu[2], Tao Wang[2], Yangyang Wei[2],

4 Shoulin Hu[1], Sumei Wan[1*], Renhai Peng[1,2*]

5 [1] College of Plant Science, Tarim University, Alar, Xinjiang, China

6 [2] Anyang Institute of Technology, Anyang, Henan, China

8 † These authors contributed equally to this work.

9 * Co-Corresponding Author:

10 Sumei Wan

11 Alar, Xinjiang, 843300, China

12 Email address: wansumei510@163.com

13 Renhai Peng

14 Anyang, Henan, 455000, China

15 Email address: aydxprh@163.com

17 Abstract

18 **Background.** Golden2-Like (GLK) transcription factors are a type of transcriptional regulator in

19 plants. They play a pivotal role in the plant physiological activity process and abiotic stress

20 response.

21 **Methods.** In this study, the potential function of *GLK* family genes in *Gossypium hirsutum* was

22 studied based on genomic identification, phylogenetic analysis, chromosome mapping and cis-

23 regulatory element prediction. Gene expression of nine key genes were analyzed by qRT-PCR

24 experiments.

25 **Results.** Herein, we identified a total of 146 *GhGLK* genes in *Gossypium hirsutum*, which were

26 unevenly distributed on each of the chromosomes. There were significant differences in the

27 number and location of genes between the At sub-genome and the Dt sub-genome. According to

28 the phylogenetic analysis, they were divided into ten subgroups, each of which had very similar

29 number and structure of exons and introns. Some cis-regulatory elements were identified through

30 promoter analysis, including five types of elements related to abiotic stress response, five types

31 of elements related to phytohormone and five types of elements involved in growth and

32 development. Based on public transcriptome data analysis, we identified nine key *GhGLKs*

33 involved in salt, cold, and drought stress. The qRT-PCR results showed that these genes had

34 different expression patterns under these stress condition, suggesting that *GhGLK* genes played

35 an important role in abiotic stress response. This study laid a theoretical foundation for the

36 screening and functional verification of genes related to stress resistance of *GLK* gene family in

37 cotton.

38 Introduction

39   Transcription Factors (TFs), also known as trans-acting factors, are known to activate or inhibit
40   the transcription of downstream target genes at appropriate times. TFs affect all aspects of plant
41   growth and development (Song et al., 2016; Ling et al., 2011). Plants are threatened by abiotic
42   stresses during their growth, including high temperature, herbicides, heavy metals, drought,
43   salinity, cold, pests and diseases (Nguyen et al., 2018). Crop yields are known to be seriously
44   threatened under abiotic stress (Riechmann et al., 2000; Wray et al., 2003). Transcription factors
45   can activate or inhibit gene transcription, affect the expression and function of their proteins, and
46   play a significant role in plant stress response and physiological activities (Chrispeels et al., 1999;
47   Mizoi et al., 2012).
48   For the first time, Golden2 (G2) has been identified that could cause maize to turn yellow
49   (Jenkins et al., 1926). Subsequently, some studies have revealed that G2 protein was an
50   important transcription factor in plant growth and development (Hall et al., 1998). Golden2-Like
51   (GLK) transcription factors are members of the *GARP* superfamily (Riechmann et al., 2000;
52   Xiao Y et al., 2019). Most *GLK* genes contain a Myb-DNA binding domain (which contains an
53   HLH motif) and a C-terminal (GCT) box (Liu et al., 2017). In addition, some members of
54   subgroups have conserved MYB-CC-LHEQLE domains (Qin et al., 2021).
55   *GLK* genes have been identified with different function in *Physcomitrella patens*,
56   *Arabidopsis thaliana*, *Oryza sativa*, *Capsicum annuum* L. and *Solanum lycopersicum* (Hall et al.,
57   1998; Rossini et al., 2001; Fitter et al., 2002; Waters et al., 2009; Yasumura et al., 2005). In
58   terms of cell differentiation, functional redundancy of the *GLKs* is present in *A. thaliana* and rice
59   (Chen et al., 2016). In corn, *ZmGLK1* and *ZmGLK2*, a pair of homologous genes, have basically
60   the same function and are expressed in mesophyll cells and vascular bundle sheath cells,
61   respectively (Chang et al., 2012). These results indicate that *GLKs* are expressed differently in
62   different photosynthetic cells, which control the process of cell differentiation and play an
63   important role in chloroplast development (Yasumura et al., 2005). Studies have indicated that
64   *GLK1* is mainly expressed in leaf tissue and *GLK2* is mainly expressed in fruit (Chen et al.,
65   2016). In tomato, *SlGLK2* is only expressed in the fruit, and changes the content of sugar and
66   carotenoid by regulating chloroplast development, thus affecting the fruit quality of tomato
67   (Powell et al., 2012; Nguyen et al., 2014). In addition, some studies have found that the
68   *KNOTTED1-LIKE HOMEOBOX (KNOX)* gene acts on the downstream of *SlGLK2* and only
69   affects the chloroplast development in tomato fruit but not in leaf tissue (Nadakuduti et al., 2014).
70   All these indicate that although *GLK1* and *GLK2* have the same function, they have different
71   regulatory pathways in different organs and have tissue-specific characteristics (Nguyen et al.,
72   2014). *GLKs* may enrich the carbon utilization rate of plants to a certain extent and promote the
73   growth and exploitation of plants by increasing the fixation of carbon dioxide in root (Kobayashi
74   et al., 2014). However, the expression of *GLKs* in leaves was significantly higher than that in
75   roots (Fitter et al., 2002).
76   The *GLKs* also play an important role in biotic stress. At present, studies on *GLKs* in plant
77   disease resistance mainly focus on *A. thaliana*, a model system, and some studies on rice, while
78   few studies are known on other crops (Chen et al., 2016). The overexpression of *AtGLK1*
79   enhances the resistance of *A. thaliana* infected with *Fusarium graminearum*, and plays a positive

80 role in the tolerance to *Cucumber Mosaic Virus* (Savitch et al., 2007; Schreiber et al., 2011; Han
81 et al., 2016). *AtGLK1* regulates genes related to disease resistance, and its effect on different
82 pathogens are different (Chen et al., 2016). In addition to *A. thaliana*, *OsGLK1* in rice has also
83 shown certain disease resistance (Nakamura et al., 2009). In addition, *GLKs* are also involved in
84 hormonal response. The resistance of *AtGLK1* to *Hyalo-peronospora arabidopsidis* Noco2
85 involves two signaling pathways: salicylic acid and jasmonic acid (Murmu et al., 2014). The
86 *GLK1/2-WRKY40* transcription module plays a negative regulatory role in the ABA response
87 (Ahmad et al., 2019).
88   Many functions of *GLK* genes have been studied in depth, but the functions related to
89 abiotic stress are rarely mentioned and few studies have been reported. Research has found that
90 some genes in the *GLK* gene family are involved in stressresponse in maize (Liu et al., 2016). It
91 was found that the down-regulation of the *SlGLK29* gene in tomato could reduce the cold
92 resistance (Liu, 2018). This suggests that different *GLK* genes have distinct expression patterns
93 under different stress condition.
94   Cotton is an important fiber and oil crop in the world, but its development is always
95 threatened by abiotic stresses, including extreme temperature, drought and salt (Zhang et al.,
96 2018; Li et al., 2021). These stresses will become more severe in the future, which will lead to
97 lower crop yields and quality (Onaga et al., 2016). Therefore, it is necessary to identify genes
98 related to abiotic stress resistance to assist cotton genetic improvement. So far, the identification
99 of the *GLK* gene family in *G. hirsutum* has not been reported. In our research, the *GLK* gene
100 family in *G. hirsutum* was identified, and the subcellular localization, chromosomal distribution,
101 gene structure and expression level of *GLK* genes were analyzed. The results demonstrate that
102 nine key genes responded to drought, salt and cold stress. This study provides a reference for
103 further study of the role of *GLKs* in the stress response of cotton.

Materials & Methods
104

Identification of the *GLK* genes in *Gossypium hirsutum*.
105
106 The genome files of *G. hirsutum* (TM-1 HAU_v1.1) and *Theobroma cacao* L. (assembly
107 Criollo_cocoa_genome_V2) were obtained from CottonFGD (http://www.cottonfgd.org/) (Wang
108 et al., 2019) and NCBI (https://www.ncbi.nlm.nih.gov/) (Argout et al., 2011). The genome and
109 protein sequences of *A. thaliana* were obtained from NCBI (https://www.ncbi.nlm.nih.gov/). The
110 Markov Model (HMM) of PF00249.31 and PF14379.6 were downloaded from Pfam
111 (https://pfam.xfam.org/). The TBtools v1.087 HMM Search (Chen et al., 2020) was used to
112 identify the *GLK* gene family in *A. thaliana, G. hirsutum* and *T. cacao* with PF00249.31 and
113 PF14379.6, which were most likely members of *GLK* gene family. Published tomato GLK
114 protein sequences were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) (Liu, 2018).
115 The *A. thaliana* and tomato sequences were used as query sequences to BLAST against the
116 cotton and *T. cacao* genome database using TBtools (Chen et al., 2020). ProtParam tool
117 (https://web.expasy.org/protparam/) was used to predict the physical and chemical properties of
118 GLK proteins, including the number of amino acids, molecular weight(MW), isoelectric
119 point(PI), and Instability index. The subcellular localization of *GLKs* was predicted through the

120   WoLF PSORT (https://wolfpsort.hgc.jp/) and CELLO v.2.5 (http://cello.life.nctu.edu.tw/)

121   resources (Yu et al., 2006).

122   **Phylogenetic analysis of *GLK* genes.**

123   A phylogenetic tree was constructed the obtained GLK protein sequences of *A. thaliana*, tomato,

124   *T. cacao* and *G. hirsutum*. A neighbor-joining tree of *GLK* genes was constructed using MEGA-

125   X with 1000 bootstrap replications (Hall, 2013). The phylogenetic tree was drawn using

126   EvolView (He et al., 2016).

127   **Analysis of the conserved motifs and gene structure.**

128   The conserved sequences of *GLK* was identified and analyzed by the MEME website

129   (http://meme-suite.org/) (Bailey et al., 2009). The optimal width was 10 to 150, and the number

130   of motifs was 26. Everything else is set was default values. TBtools (Chen et al., 2020) was used

131   to map conserved motifs and gene structures (introns and exons), using MAST profiles from the

132   MEME website and the GFF3 profiles for each gene.

133   **Chromosomal localization analysis of *GLK* genes.**

134   Based on the genome and genome annotation files of *G. hirsutum*, the chromosome distribution

135   of *GLK* genes and their physical locations were obtained through TBtools (Chen et al., 2020).

136   **Cis-regulatory elements analysis of *GLK*.**

137   In order to analyze the promoter of *GLK* in *G. hirsutum* and predict the function of the *GLK*

138   genes, 2000 bp sequence upstream of the start codon for each gene was extracted and input into

139   the Plantcare website for analysis (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/).

140   **GO and KEGG enrichment analysis of *GLK* genes.**

141   For functional enrichment analysis, *GhGLKs* were submitted to the omicshare tool

142   (https://www.omicshare.com/tools/) for GO and KEGG enrichment analysis.

143   **Differential gene expression analysis.**

144   In order to study the expression pattern of the *GLK* in *G. hirsutum*, the transcriptome sequencing

145   data of *G. hirsutum* (PRJNA490626) under cold, salt, and drought stress was downloaded from

146   NCBI (https://www.ncbi.nlm.nih.gov/). Trimmomatic (Bolger et al., 2014) was used to remove

147   the adapter and perform quality control. Reads were mapped to the genome using the hisat2

148   program (Kim et al., 2015), and then Fragments Per Kilobase of transcript per Million fragments

149   (FPKM) values of *GLK* genes were calculated by Cufflinks (Ghosh et al., 2016; Pollier et al.,

150   2013). The expression level of *GLK* family genes was standardized, expressed as a FPKM value,

151   and transformed in $\log_2$ form, and the heatmap was drawn using TBtools software (Chen et al.,

152   2020).

153   **Stress treatments and qRT-PCR analysis.**

154   *G. hirsutum* acc. TM-1 planted in the experimental field of Anyang Institute of Technology was

155   selected as experimental material. The seedlings with stable growth of *G. hirsutum* were treated

156   with three stress: salt stress, drought stress and cold stress. The root of cotton seedlings was

157   irrigated with 250 mM NaCl solution and 18% PEG solution to simulate salt stress and drought

158   stress. It was placed in a 4 °C incubator to simulate cold stress treatment. Leaves from cotton

159   seedlings with consistent growth were collected after 0 h, 1 h, 3 h, 6 h, 12 h, 24 h of above

160   stresses. All samples were immediately frozen in liquid nitrogen and stored at −80 °C for RNA

161 extraction.

162   Total RNA was extracted from each sample using the EASYspin Plus Plant RNA Kit
163 (RN38, Aidlab Biotech, Beijing, China). The quality of RNA was determined by agarose gel
164 electrophoresis and a Nanodrop2000 nucleic acid analyzer. The cDNAs were synthesized using a
165 TranScript All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (Transgen Biotech,
166 Beijing, China). The kit used for Real-time PCR is the TransStart Top Green qPCR SuperMix kit
167 (Transgen Biotech, Beijing, China). The instrument used is the ABI 7500 Fast Real-Time PCR
168 system (Applied Biosystems, USA). The specific primers for these differentially expressed genes
169 were designed using the Prime-Blast in the NCBI online database and are listed in
170 Supplementary Table S1. Each experiment was repeated three times, and choose two groups of
171 good data to graph. The relative gene expression levels were analyzed using the $2^{-\Delta\Delta Ct}$ method
172 (Livak ＆ Schmittgen, 2001).

Results

173

174 **Identification and analysis of basic physicochemical properties of *GLK* family members in**
175 ***G. hirsutum*.**

176 A total of 146 *GLK* genes were identified from *G. hirsutum*, named *GhGLK1-GhGLK146*. The
177 basic physical and chemical properties were also predicted and analyzed. Amino acid sequences
178 range in length is from 140 (*GhGLK27*) to 826 (*GhGLK139*), the isoelectric point is from 5.07
179 (*GhGLK9*) to 9.71 (*GhGLK134*). The instability index refers to how stable the protein is in the
180 test tube (≤40, possibly stable; >40, possibly unstable). Prediction showed that, except for 13
181 genes, all other genes may be stable. According to the results of subcellular localization, all the
182 *GhGLK* genes were located in the nucleus. *GhGLK6*, *GhGLK46*, *GhGLK76*, *GhGLK128* and
183 *GhGLK139* were located in the chloroplast and nucleus. *GhGLK112* and *GhGLK127* were in
184 cytoplasm and nucleus (Table S2).

185 **Phylogenetic analysis of *GLK* in *G. hirsutum*.**
186 Using tomato, *A. thaliana* and *T. cacao* GLK protein sequences as reference, a rootless
187 phylogenetic tree of GLK protein in *G. hirsutum* was constructed (Fig. 1). As the results shown,
188 146 *GLK* genes of *G. hirsutum* were divided in ten subfamilies. At the same time, the number of
189 *GLKs* in *G. hirsutum* was much higher than that in *A. thaliana*, *T. cacao* and tomato.

190 **Gene structure and protein conserved motif analysis of *GLK* in *G. hirsutum*.**
191 The motif of each protein, namely the conserved element, was analyzed by MEME. A total of 26
192 possible motifs were identified. Motif 1 was included in all *GhGLK*s , and it was conservative for
193 *GhGLK*s. These motifs differ between subfamilies, but were conserved within each subfamily.
194 For example, motifs 17 and 18 were found only in subfamily D and E, respectively. Some
195 subfamilies had very conservative motifs. For example, all members of the C subfamily had
196 motif 5, 3, 8, 2, 1, 14, 10. Predictions about exons and introns of the *GLK* genes were shown in
197 the Fig. 2. The number of exons was at least one and at most eleven. Except for *GhGLK87*, the
198 other members of the D subfamily had one exon, and all members of the C subfamily had 11
199 exons. Most of the other members had between 4 and 7 introns. The number of exons and introns
200 varied among different subfamilies, but in the same subfamily, the exon-intron structure of most

201 members showed great similarity.

202 **Chromosomal localization analysis of *GLK* family genes in *G. hirsutum*.**

203 Based on a physical map of the *GLK* family members of *G. hirsutum*, 144 of 146 *GLK* genes
204 were located on 26 chromosomes (Fig. 3). The other two genes were located on the scaffold.
205 Among the chromosome-located genes, seventeen genes were located on chromosome 5. There
206 were sixteen genes located on chromosomes 8 and 12, only six genes were located on
207 chromosome 2 (Fig. 4). There were significant differences in At sub-genome and Dt sub-genome
208 about the number and location of genes. For example, there were two genes on chromosome A02
209 and four genes on chromosome D02. This may be due to the fact that *G. hirsutum* is a tetraploid
210 cotton species, which is a hybrid between an A-genome-like *Gossypium herbaceum* and a D-
211 genome-like *Gossypium raimondii*.

212 **Analysis of promoters of *GLK* genes.**

213 The Plantcare was used to analyze the sequence of 2000 bp upstream of the promoter region of
214 *GhGLK* genes, and the found cis-regulatory elements were shown in Fig. 5. We found several
215 cis-regulatory elements in stress response, which were associated with anaerobic induction,
216 defense and stress response, drought, cold, and wound, respectively. Meanwhile, we also found a
217 number of hormone-responsive cis-regulatory elements, which were associated with abscisic
218 acid (ABA), auxin (IAA), gibberellin (GA), salicylic acid (SA), and methyl jasmonate (MeJA).
219 Among all cis-regulatory elements, the number of elements related to light response was the
220 largest and the distribution was the widest. In addition, we had identified cis-regulatory elements
221 involved in cell cycle regulation, circadian rhythm, down regulation of photosin expression, and
222 regulation of flavonoid biosynthesis genes. In summary, different types and quantities of cis-
223 regulatory elements were distributed in different *GhGLK* promoters of cotton. According to the
224 results, it was speculated that under environmental stress, the cis-regulatory elements leaded to
225 the expression of *GhGLK* genes, thereby enhancing the resistance to environmental stress.

226 **GO and KEGG enrichment analysis of *GLK* in *G. hirsutum*.**

227 In order to further understand the function of *GhGLKs*, we carried out functional enrichment
228 annotation of gene ontology (GO) and kyoto encyclopedia of genes and genomes (KEGG). The
229 results improved our accurate understanding of gene function, including many significantly
230 enriched terms (Fig. 6, Table S2). The GO enrichment analysis of *GhGLKs* was divided into 3
231 categories, including biological processes, molecular functions and cellular components, of
232 which the biological process had the largest number of 13 GO entries, followed by molecular
233 functions and cellular components. The biological process was mainly concentrated in four
234 subclasses: cellular process, metabolic process, regulation of biological process and biological
235 regulation. The molecular function mainly included binding, nucleic acid binding transcription
236 factor activity and catalytic activity. The cellular components were mainly organelle subclasses.
237 KEGG pathway enrichment analysis of *GLK* family genes revealed two pathways, RNA
238 transport and plant hormone signal transduction respectively. To sum up, the functional
239 enrichment analysis results confirmed the functions of *GhGLKs* in many biological processes,
240 which were related to plant growth and development, plant hormone signal transduction and so
241 on.

242 **Expression analysis of *GLK* gene in *G. hirsutum* under different abiotic stresses.**

243 To investigate the response of *GhGLK* to adversity stress, we analyzed the expression levels of
244 146 *GhGLKs* under different adversity conditions. The results showed that the expression of
245 *GhGLK* genes altered under salt, drought and cold stress, which revealed that *GhGLKs* were
246 involved in the regulation of adversity stress (Fig. 7). According to the above results, we selected
247 nine key genes with differentially expressed *GhGLK*s and analyzed the expression pattern by
248 qRT-PCR (Fig. 8).

249 Under salt stress, the expression levels of all *GhGLK* genes increased after 1 h, reached a
250 minimum at 12 h, and increased again at 24 h. Among them, the expression levels of *GhGLK38*,
251 *GhGLK55*, *GhGLK82*, and *GhGLK120* increased significantly at a certain moment, reaching the
252 highest level at 1 h, 24 h, 6 h, and 1 h respectively. Except for *GhGLK1*, *GhGLK39* and
253 *GhGLK82*, other genes increased after 1 h of treatment, then continued to decline, and increased
254 again after 24 h. These genes had a certain trend under salt stress treatment, and the response to
255 salt stress was more obvious, and the expression of genes also had certain differences.

256 Under drought stress, *GhGLK1*, *GhGLK8*, and *GhGLK120* genes were significantly down-
257 regulated after 1 h of treatment, and their expression levels were extremely low, indicating that
258 these genes may play a negative regulatory role under drought stress. *GhGLK46* and *GhGLK55*
259 genes changed significantly, both were up-regulated after 1 h of treatment and reached a peak at
260 3 h, decreased after 6 h, and then continued to increase. There was no significant change in the
261 expression levels of the two genes *GhGLK39* and *GhGLK47*.

262 Under cold stress, all *GhGLK* genes were significantly lower than that under drought stress
263 and salt stress, which indicated that *GhGLK* genes were not strongly induced by low temperature
264 stress. Except for *GhGLK8*, *GhGLK38*, *GhGLK82* and *GhGLK120*, the other genes were up-
265 regulated after 1 h of low temperature treatment. *GhGLK55* was strongly induced by cold stress,
266 and its expression continued to increase, reaching a peak at 3 h and beginning to decrease.
267 *GhGLK38* reached the lowest after 12 h of cold stress treatment, *GhGLK39* and *GhGLK82*
268 reached the lowest at 3 h and 1 h, respectively, and all other *GhGLKs* reached the lowest at 24 h.

Discussion

270 As the main cultivated cotton species in the world, *G. hirsutum* is extremely important to human
271 life. Adversity coercions, including drought, salt and cold, are a very serious problem to plant
272 growth and development. Golden2-Like is a type of transcription factor that exists widely in
273 plants, and belongs to an important type of transcription factor in the GARP transcription factor
274 superfamily of Myb transcription factors. It had been identified in many species and found to be
275 closely related to abiotic stress, but it had not been reported in cotton.

276 Herein, a total of 146 *GLKs* were identified from *G. hirsutum*, named each gene according
277 to its position on the chromosome. The number of *GLK* gene in cotton was much higher than that
278 in tomato (54 genes) (Liu, 2018) and maize (59 genes) (Liu et al., 2016), which might be related
279 to the relatively large genome of *G. hirsutum* heterotetraploid and the complexity of gene
280 regulationdue to the complexity of the genome. The analysis of the physical and chemical
281 properties of upland cotton *GhGLK* showed that its sequence length, relative molecular weight

282   and isoelectric point distribution range were very large. This might be due to the large-scale

283   replication of the upland cotton genome and the large number of upland cotton *GhGLK* genes.

284   This study predicted the location of *G. hirsutum GLK*s in the cell, and found that some genes

285   were located in the chloroplast. Previous studies had shown that the *GLK* gene is related to the

286   growth and development of the chloroplast. This result was also consistent with the results of

287   previous studies.

288   We analyzed the phylogenetic relationship of *G. hirsutum GLKs*, and constructed a

289   phylogenetic tree of tomato, *A. thaliana* and *G. hirsutum*. The results showed that the *GLK* gene

290   of *G. hirsutum* can be divided into 11 subfamilies, while tomato and corn were divided into 6

291   subfamilies (Liu, 2018). This might be due to the fact that the number of *GLK* genes in cotton

292   was larger than that of *GLK* in tomato, resulting in a more precise and accurate sub-family

293   classification. Some *GhGLK* and tomato *GLK* were homologous, and the results of this study

294   indicated that there was a close evolutionary relationship between *GLK* genes. At the same time,

295   we found that *GhGLK* gathered in the same subfamily had similar motif distribution patterns,

296   motif positions and lengths.

297   The gene structure analysis also showed that the *GLK* gene had strong evolutionary

298   conservation. *GLK* genes in the same subgroup had similar number of exons/introns and length

299   of exons. The results of previous studies showed that the structure of the exons and introns of

300   each subfamily of *GLK* genes in tomato showed a large similarity, and the number of introns of

301   most members was between 4 and 7. The results were the same in *G. hirsutum*, except that there

302   are up to 11 exons in upland cotton and up to 7 in tomato(Liu, 2018).

303   In the *G. hirsutum GLK* gene family, 144 genes were located on 26 chromosomes of group

304   A and D, and the other 2 genes were located on scaffolds of unknown chromosomes. Among

305   them, there were 70 *GLK* genes on chromosomes of group A and chromosomes of group D.

306   There were 74 *GLK* genes on it. The distribution of the *GLK* gene in *G. hirsutum* was uneven on

307   the chromosomes, but the distribution of the *GLK* gene on the two homologous chromosomes

308   was indeed the same.

309   Many cis-elements in *GhGLK* promoter were related to biotic and abiotic stress. In order to

310   further understand the function of *GhGLK* under different environmental stresses, we analyzed

311   the expression patterns of 9 *GhGLK* under different environmental stresses by qRT-PCR. This

312   showed that most selected *GhGLKs* respond to abiotic stresses, including drought, cold, and salt.

313   In previous studies, it was found that the expression pattern of *ZmGLK3* gene in maize was up-

314   regulated under drought and salt stress (Liu et al., 2016); the expression of *SlGLK7* in tomato

315   was up-regulated under three stresses (Liu, 2018). The expression of *SlGLK15* and *SlGLK37*

316   genes in tomato decreased first and then increased during cold stress and salt stress, and these

317   two genes belonged to the same subfamily(Liu, 2018). In our results, we found that the two

318   genes *GhGLK55* and *GhGLK120* were located in the same subgroup, and they were in the same

319   subgroup as *SlGLK15* and *SlGLK37* in tomato. These two genes in *G. hirsutum* also decreased

320   firstly and then increased under cold and salt stress, indicating that the orthologous genes they

321   should have similar expression patterns in the face of adversity and coercion. We also found that

322   the differentially expressed genes all have anaerobic-induced and light-responsive cis-acting

323 elements, and they all contained at least one abiotic stress-responsive element. *GhGLK120* had
324 the most hormone-responsive elements.

Conclusions

326 In this study, 146 *GhGLK* genes were identified in *G. hirsutum*. Based on the analysis of their
327 physical and chemical properties, subcellular localization, phylogenetic relationship, distribution
328 of cis-regulatory elements, we identified nine key *GhGLKs* involved in salt, cold, and drought
329 stress. The qRT-PCR results of these genes in three stress responses showed that *GhGLKs* played
330 an important role in abiotic stress responses in cotton. It was of great significance to make full
331 use of its resistance germplasm resources and provided a theoretical basis for further mining of
332 cotton resistance genes.

References

334 **Ahmad R, Liu Y, Wang TJ, Meng Q, Yin H, Wang X, Wu Y, Nan N, Liu B, Xu ZY. 2019.**
335 GOLDEN2-LIKE Transcription Factors Regulate WRKY40 Expression in Response to
336 Abscisic Acid. Plant Physiology 179(4):1844-1860. DOI 10.1104/pp.18.01466.
337 **Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C,**
338 **Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y,**
339 **Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JS, Schuster SC,**
340 **Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Bérard A, Viot C,**
341 **Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma Z, Zhang Y, Brown S,**
342 **Bourge M, Golser W, Song X, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B,**
343 **Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR,**
344 **Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C. 2011.** The genome of
345 Theobroma cacao. Nature Genetics **43(2)**:101-8. DOI 10.1038/ng.736.
346 **Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.**
347 **2009.** MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research
348 **37(Web Server issue)**:W202-8. DOI 10.1093/nar/gkp335.
349 **Bolger AM, Lohse M, Usadel B. 2014.** Trimmomatic: a flexible trimmer for Illumina sequence
350 data. Bioinformatics **30(15)**:2114-20. DOI 10.1093/bioinformatics/btu170.
351 **Chang YM, Liu WY, Shih AC, Shen MN, Lu CH, Lu MY, Yang HW, Wang TY, Chen SC,**
352 **Chen SM, Li WH, Ku MS. 2012.** Characterizing regulatory and functional differentiation
353 between maize mesophyll and bundle sheath cells by transcriptomic analysis. Plant Physiology
354 **160(1)**:165-77. DOI 10.1104/pp.112.203810.
355 **Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020.** TBtools: An
356 Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Molecular
357 Plant **13(8)**:1194-1202. DOI 10.1016/j.molp.2020.06.009.
358 **Chen M, Ji M, Wen B, Liu L, Li S, Chen X, Gao D, Li L. 2016.** GOLDEN 2-LIKE
359 Transcription Factors of Plants. Frontiers in Plant Science **7**:1509. DOI
360 10.3389/fpls.2016.01509.

361 **Chrispeels HE, Oettinger H, Janvier N, Tague BW. 2000.** AtZFP1, encoding Arabidopsis
362   thaliana C2H2 zinc-finger protein 1, is expressed downstream of photomorphogenic activation.
363   Plant Molecular Biology **42(2)**:279-90. DOI 10.1023/a:1006352809700.
364 **Fitter DW, Martin DJ, Copley MJ, Scotland RW, Langdale JA. 2002.** *GLK* gene pairs
365   regulate chloroplast development in diverse plant species. The Plant Journal **31(6)**:713-27.
366   DOI 10.1046/j.1365-313x.2002.01390.x.
367 **Ghosh S, Chan CK. 2016.** Analysis of RNA-Seq Data Using TopHat and Cufflinks. Methods in
368   molecular biology **1374**:339-61. DOI 10.1007/978-1-4939-3167-5_18.
369 **Hall LN, Rossini L, Cribb L, Langdale JA. 1998.** GOLDEN 2: a novel transcriptional
370   regulator of cellular differentiation in the maize leaf. Plant Cell **10(6)**:925-36. DOI
371   10.1105/tpc.10.6.925.
372 **Hall BG. 2013.** Building phylogenetic trees from molecular data with MEGA. Molecular
373   Biology Evolution **30(5)**:1229-35. DOI 10.1093/molbev/mst012.
374 **Han XY, Li PX, Zou LJ, Tan WR, Zheng T, Zhang DW, Lin HH. 2016.** GOLDEN2-LIKE
375   transcription factors coordinate the tolerance to Cucumber mosaic virus in Arabidopsis.
376   Biochemical and Biophysical Research Communications **477(4)**:626-632. DOI
377   10.1016/j.bbrc.2016.06.110.
378 **He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. 2016.** Evolview v2: an online
379   visualization and management tool for customized and annotated phylogenetic trees. Nucleic
380   Acids Research **44(W1)**:W236-41. DOI 10.1093/nar/gkw370.
381 **Jenkins M.T. 1926.** A second gene producing golden plant color in maize. The American
382   Naturalist **60(670)**:484-488.
383 **Kim D, Langmead B, Salzberg SL. 2015.** HISAT: a fast spliced aligner with low memory
384   requirements. Nature Methods **12(4)**:357-60. DOI 10.1038/nmeth.3317.
385 **Kobayashi K, Baba S, Obayashi T, Sato M, Toyooka K, Keränen M, Aro EM, Fukaki H,**
386   **Ohta H, Sugimoto K, Masuda T. 2012.** Regulation of root greening by light and
387   auxin/cytokinin signaling in *Arabidopsis*. Plant Cell **24(3)**:1081-95. DOI
388   10.1105/tpc.111.092254.
389 **Li Z, Liu Z, Wei Y, Liu Y, Xing L, Liu M, Li P, Lu Q, Peng R. 2021.** Genome-wide
390   identification of the MIOX gene family and their expression profile in cotton development and
391   response to abiotic stress. PLoS One **16(7)**:e0254111. DOI 10.1371/journal.pone.0254111.
392 **Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, Huang S, Xie B. 2011.** Genome-wide
393   analysis of *WRKY* gene family in *Cucumis sativus*. BMC Genomics **12**:471. DOI
394   10.1186/1471-2164-12-471.
395 **Liu F, Xu Y, Han G, Zhou L, Ali A, Zhu S, Li X. 2016.** Molecular Evolution and Genetic
396   Variation of *G2-Like* Transcription Factor Genes in Maize. PLoS One **11(8)**:e0161763. DOI
397   10.1371/journal.pone.0161763.
398 **Liu G, Zhao TT, Yang HH, Zhang DY, Jiang JB, Li JF and Xu XY. 2016.** Research progress
399   on transcriptomics in tomato. Jiyinzuxue Yu Yingyong Shengwuxue (Genomics and Applied
400   Biology) **35(10)**:2802-2807.

401 **Liu JF. 2018.** Bioinformatics analysis of tomato G2-like transcription factor family and
402   identification of resistance-related genes. D. Agriculture. Thesis. College of Horticulture &
403   Landscape Architecture.

404 **Liu JF, Zhang J, Li H, Zhao TT, and Li JF, 2017**, Research progress of plant GOLDEN2-
405   Like transcription factor, Fenzi Zhiwu Yuzhong (Molecular Plant Breeding) **15(10)**: 3949-
406   3956.

407 **Livak KJ, Schmittgen TD. 2001.** Analysis of relative gene expression data using real-time
408   quantitative PCR and the 2 (-Delta Delta C(T)) Method. Methods **25(4)**:402-8. DOI
409   10.1006/meth.2001.1262.

410 **Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012.** AP2/ERF family transcription factors in
411   plant abiotic stress responses. Biochim Biophys Acta **1819(2)**:86-96. DOI
412   10.1016/j.bbagrm.2011.08.004.

413 **Murmu J, Wilton M, Allard G, Pandeya R, Desveaux D, Singh J, Subramaniam R. 2014.**
414   Arabidopsis GOLDEN2-LIKE (GLK) transcription factors activate jasmonic acid (JA)-
415   dependent disease susceptibility to the biotrophic pathogen Hyaloperonospora arabidopsidis,
416   as well as JA-independent plant immunity against the necrotrophic pathogen Botrytis cinerea.
417   Molecular Plant Pathology **15(2)**:174-84. DOI 10.1111/mpp.12077.

418 **Nadakuduti SS, Holdsworth WL, Klein CL, Barry CS. 2014.** *KNOX* genes influence a
419   gradient of fruit chloroplast development through regulation of GOLDEN2-LIKE expression
420   in tomato. The Plant Journal **78(6)**:1022-33. DOI 10.1111/tpj.12529.

421 **Nakamura H, Muramatsu M, Hakata M, Ueno O, Nagamura Y, Hirochika H, Takano M,**
422   **Ichikawa H. 2009.** Ectopic overexpression of the transcription factor *OsGLK1* induces
423   chloroplast development in non-green rice cells. Plant Cell Physiology **50(11)**:1933-49. DOI
424   10.1093/pcp/pcp138.

425 **Nguyen CV, Vrebalov JT, Gapper NE, Zheng Y, Zhong S, Fei Z, Giovannoni JJ. 2014.**
426   Tomato GOLDEN2-LIKE transcription factors reveal molecular gradients that function during
427   fruit development and ripening. Plant Cell **26(2)**:585-601. DOI 10.1105/tpc.113.118794.

428 **Nguyen HC, Lin KH, Ho SL, Chiang CM, Yang CM. 2018.** Enhancing the abiotic stress
429   tolerance of plants: from chemical treatment to biotechnological approaches. Plant Physiology
430   **164(4)**:452-466. DOI 10.1111/ppl.12812.

431 **Onaga G, Wydra K. 2016.** Advances in Plant Tolerance to Abiotic Stresses. Plant Genomics
432   http://dx.doi.org/10.5772/64350.

433 **Pollier J, Rombauts S, Goossens A. 2013.** Analysis of RNA-Seq data with TopHat and
434   Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures.
435   Methods in molecular biology **1011**:305-15. DOI 10.1007/978-1-62703-414-2_24.

436 **Powell AL, Nguyen CV, Hill T, Cheng KL, Figueroa-Balderas R, Aktas H, Ashrafi H, Pons**
437   **C, Fernández-Muñoz R, Vicente A, Lopez-Baltazar J, Barry CS, Liu Y, Chetelat R,**
438   **Granell A, Van Deynze A, Giovannoni JJ, Bennett AB. 2012.** Uniform ripening encodes a
439   Golden 2-like transcription factor regulating tomato fruit chloroplast development. Science
440   **336(6089)**:1711-5. DOI 10.1126/science.1222218.

441 **Qin M, Zhang B, Gu G, Yuan J, Yang X, Yang J, Xie X. 2021.** Genome-Wide Analysis of the
442   G2-Like Transcription Factor Genes and Their Expression in Different Senescence Stages of
443   Tobacco (Nicotiana tabacum L.). Frontiers in Genetics 31(12):626352. DOI
444   10.3389/fgene.2021.626352.

445 **Rossini L, Cribb L, Martin DJ, Langdale JA. 2001.** The maize *golden2* gene defines a novel
446   class of transcriptional regulators in plants. Plant Cell **13(5)**:1231-44. DOI
447   10.1105/tpc.13.5.1231.

448 **Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O,**
449   **Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D,**
450   **Sherman BK, Yu G. 2000.** *Arabidopsis* transcription factors: genome-wide comparative
451   analysis among eukaryotes. Science **290(5499)**:2105-10. DOI 10.1126/science.290.5499.2105.

452 **Savitch LV, Subramaniam R, Allard GC, Singh J. 2007.** The *GLK1* 'regulon' encodes disease
453   defense related proteins and confers resistance to *Fusarium graminearum* in *Arabidopsis*.
454   Biochemical and Biophysical Research Communications **359(2)**:234-8. DOI
455   10.1016/j.bbrc.2007.05.084.

456 **Schreiber KJ, Nasmith CG, Allard G, Singh J, Subramaniam R, Desveaux D. 2011.** Found
457   in translation: high-throughput chemical screening in *Arabidopsis thaliana* identifies small
458   molecules that reduce Fusarium head blight disease in wheat. Mol Plant Microbe Interact
459   **24(6)**:640-8. DOI 10.1094/MPMI-09-10-0210.

460 **Song A, Wu D, Fan Q, Tian C, Chen S, Guan Z, Xin J, Zhao K, Chen F. 2016.**
461   Transcriptome-wide identification and expression profiling analysis of chrysanthemum trihelix
462   transcription factors. International Journal of Molecular Sciences **17(2)**:198. DOI
463   10.3390/ijms17020198.

464 **Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, Liu F, Pei L, Wang P, Zhao G, Ye Z, Huang**
465   **H, Yan F, Ma Y, Zhang L, Liu M, You J, Yang Y, Liu Z, Huang F, Li B, Qiu P, Zhang Q,**
466   **Zhu L, Jin S, Yang X, Min L, Li G, Chen LL, Zheng H, Lindsey K, Lin Z, Udall JA,**
467   **Zhang X. 2019.** Reference genome sequences of two cultivated allotetraploid cottons,
468   *Gossypium hirsutum* and *Gossypium barbadense*. Nature Genetics **51(2)**: 224-229. DOI
469   10.1038/s41588-018-0282-x.

470 **Waters MT, Wang P, Korkaric M, Capper RG, Saunders NJ, Langdale JA. 2009.** GLK
471   transcription factors coordinate expression of the photosynthetic apparatus in *Arabidopsis*.
472   Plant Cell **21(4)**:1109-28. DOI 10.1105/tpc.108.065250.

473 **Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA. 2003.**
474   The evolution of transcriptional regulation in eukaryotes. Molecular Biology and Evolution
475   **20(9)**:1377-419. DOI 10.1093/molbev/msg140.

476 **Xiao Y, You S, Kong W, Tang Q, Bai W, Cai Y, Zheng H, Wang C, Jiang L, Wang C, Zhao**
477   **Z, Wan J. 2019.** A GARP transcription factor anther dehiscence defected 1 (OsADD1)
478   regulates rice anther dehiscence. Plant Molecular Biology **101(4-5)**: 403-414. DOI
479   10.1007/s11103-019-00911-0.

480 **Yu CS, Chen YC, Lu CH, Hwang JK. 2006.** Prediction of protein subcellular localization.
481   Proteins **64(3)**:643-51. DOI 10.1002/prot.21018.

482   **Zhang B, Liu J, Yang ZE, Chen EY, Zhang CJ, Zhang XY, Li FG. 2018.** Genome-wide
483   analysis of *GRAS* transcription factor gene family in *Gossypium hirsutum L.* BMC Genomics
484   **19(1)**:348. DOI 10.1186/s12864-018-4722-x.

Figure 1

Phylogenetic analysis of *GLK* proteins from *Gossypium hirsutum*, *Theobroma cacao* L., tomato and *Arabidopsis thaliana*.

The blue marks are members of the tomato *GLK* family; The green circles are members of the *A. thaliana GLK* family; The purple squares are members of the *T. cacao GLK* family; The yellow star is the *GLK* member of cotton subgroup A, and the red triangle is the D subgroup. The blue squares are *GLK* genes located on the scaffolds.

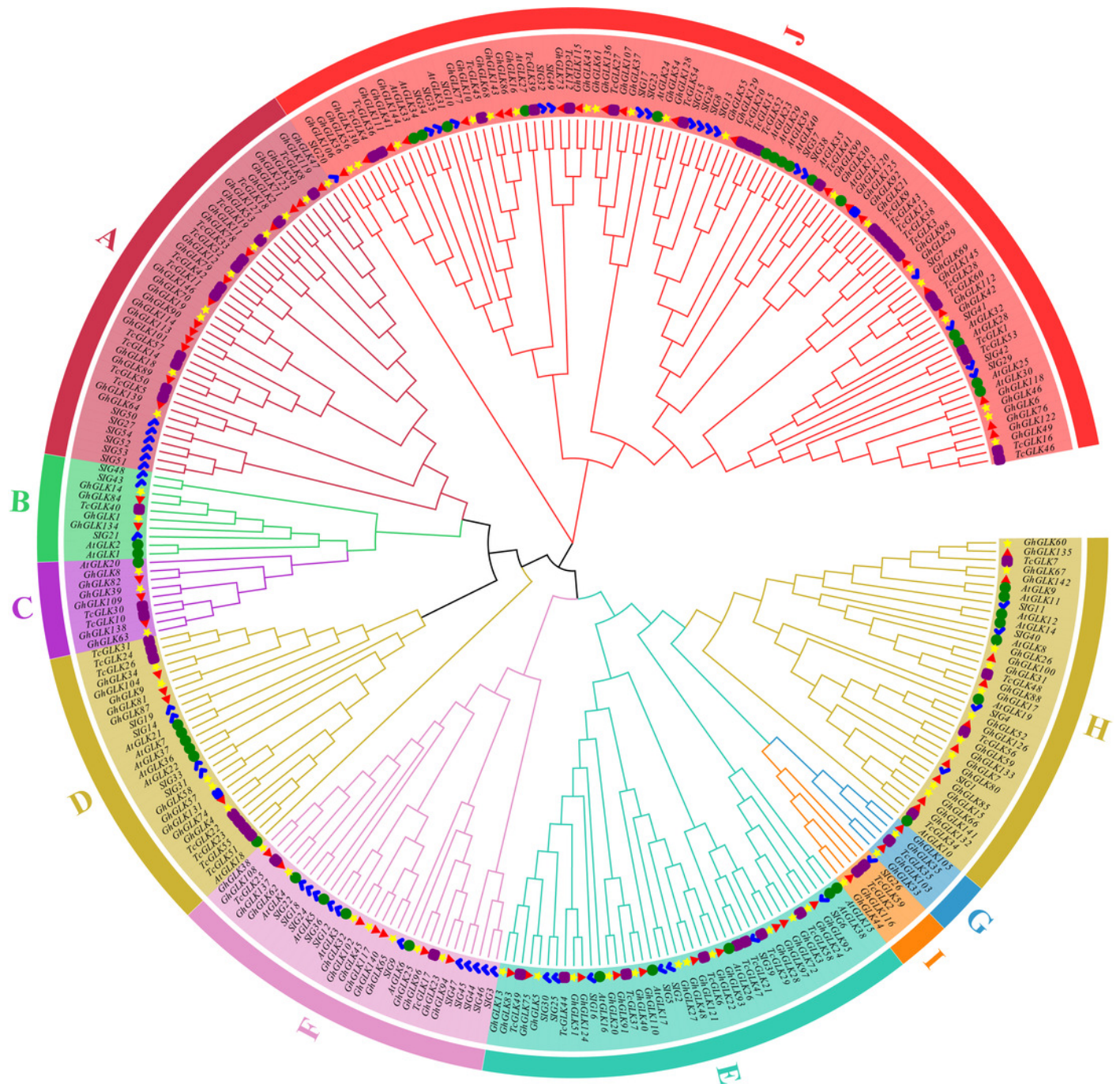



Figure 2

The conserved motif and exon-intron structure of *GLK* genes in *Gossypium hirsutum.*

(a)Phylogenetic analysis of *GhGLK* genes. (b)Analysis of conserved motif of *GhGLK* protein sequences. Different motifs are shown in a specific color. (c) Intron and exon analysis of *GhGLK* genes. Exons and introns are represented by yellow boxes and thin lines, respectively. The UTR is shown in a green box.

Figure 3

Chromosome distribution statistics of *GhGLK* genes.

Blue represents the number of genes locate on each chromosome of subgroup A, while red represents the number of genes locate on each chromosome of subgroup D.

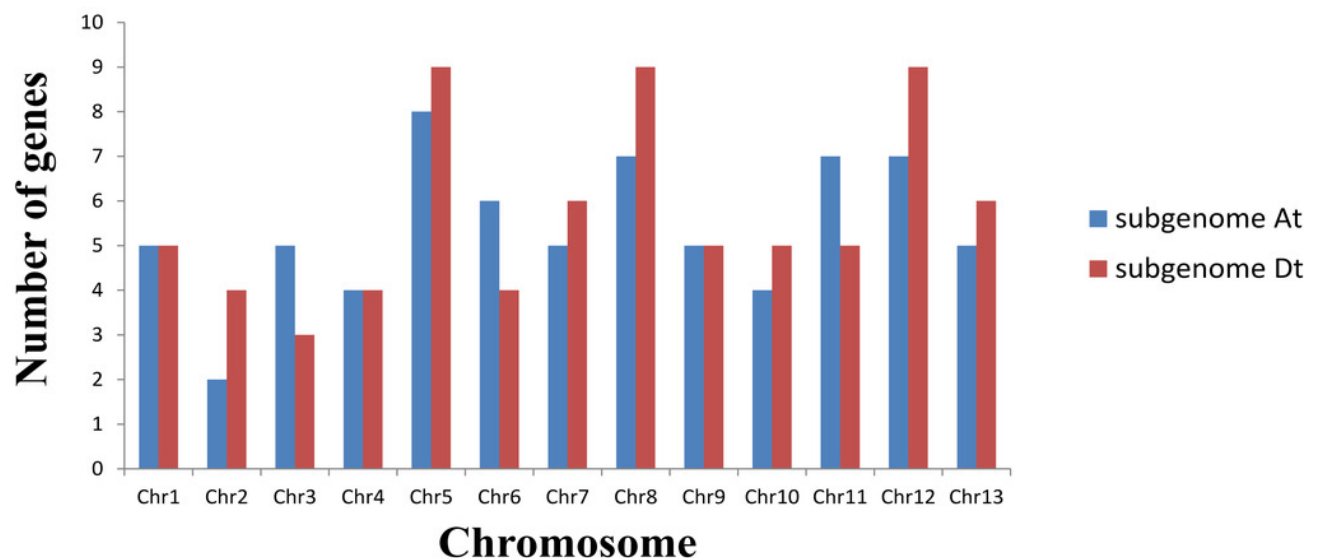



Figure 4

Location of *GhGLK* genes on chromosome.

GhGLKs are located on 26 chromosomes of *Gossypium hirsutum*, and two genes are located on scaffold. Chromosome names are shown on the left and gene names are shown on the right.

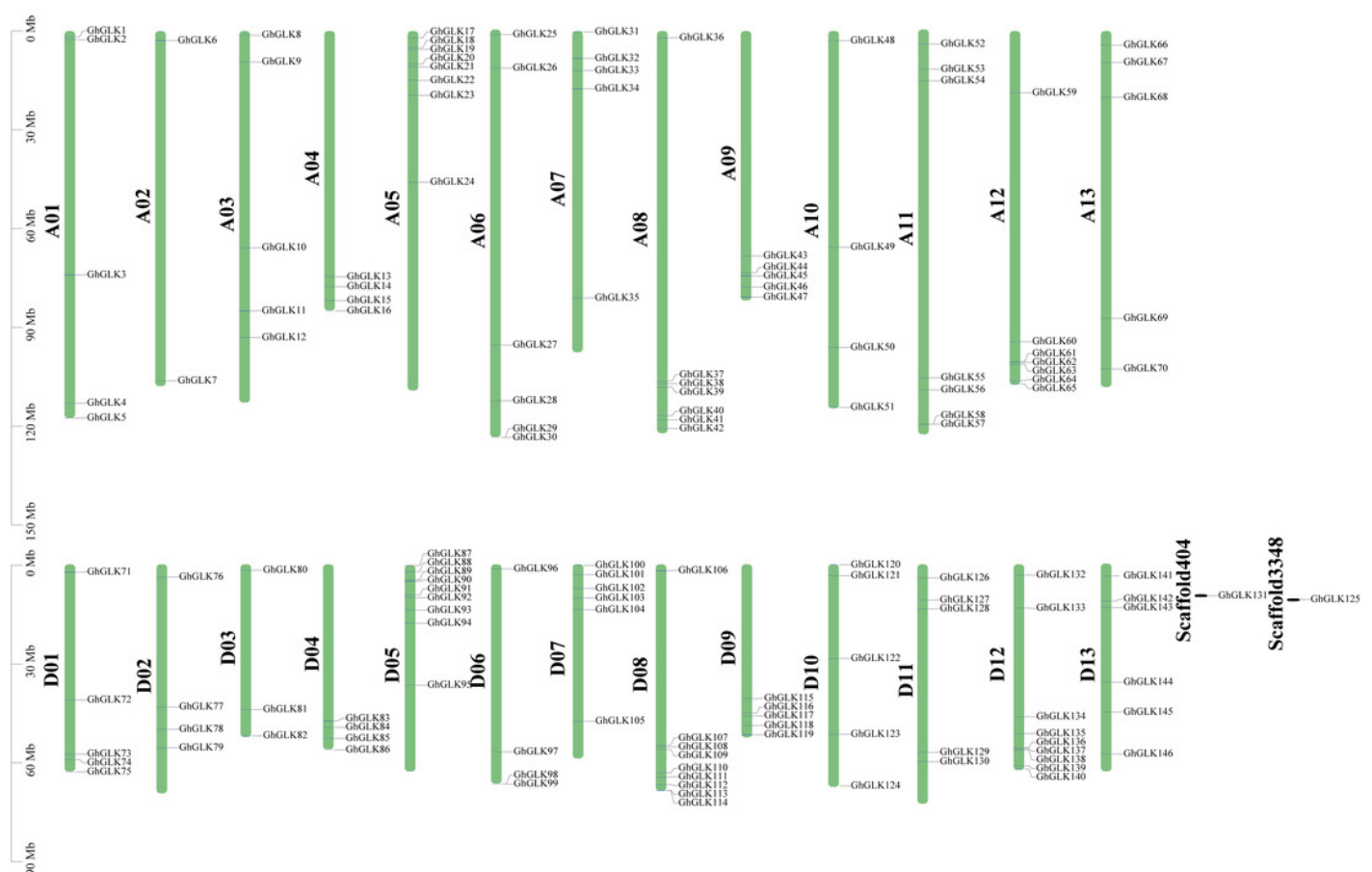



Figure 5

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

Cis-regulatory elements with the same function are shown in the same color.

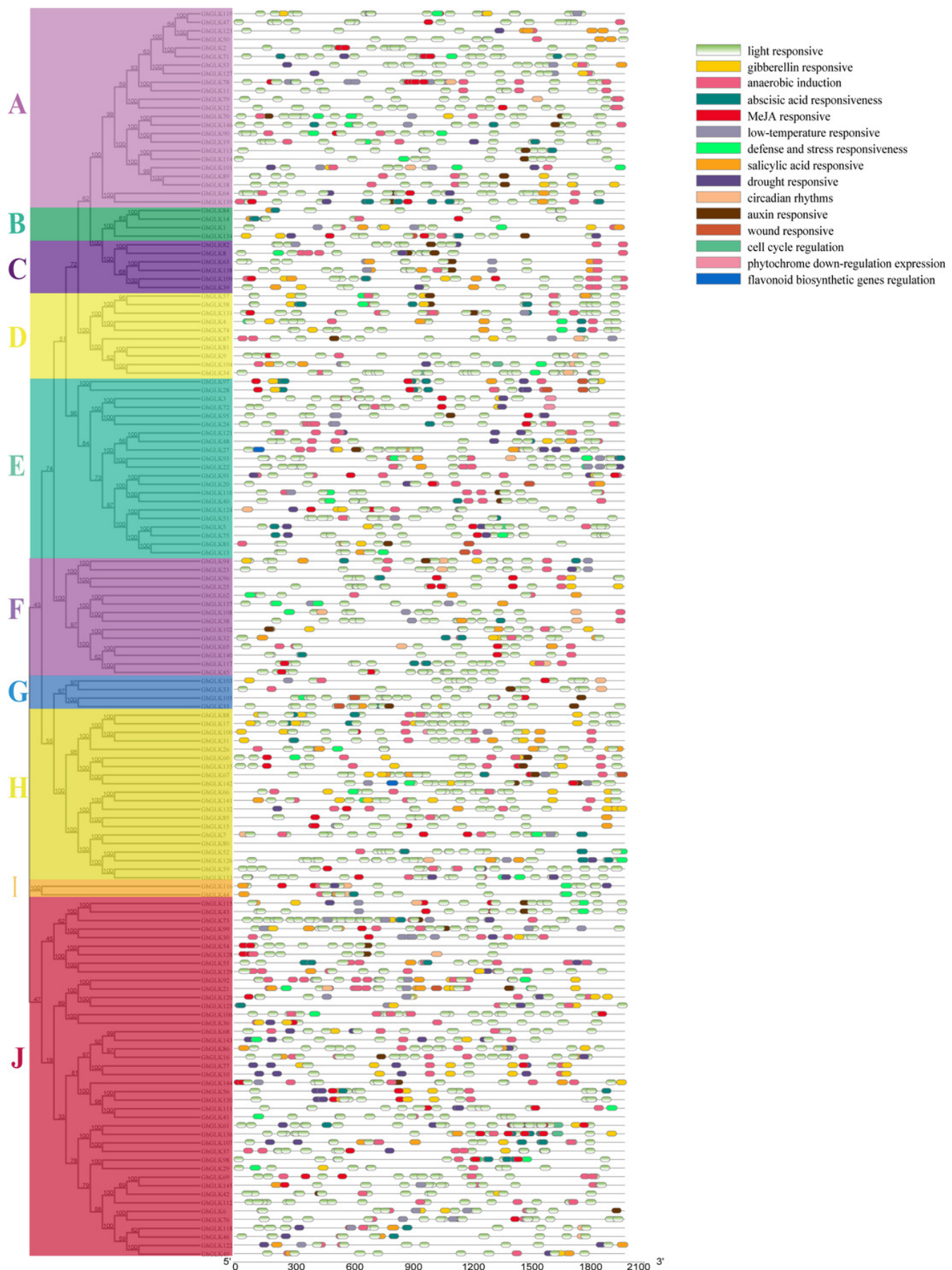
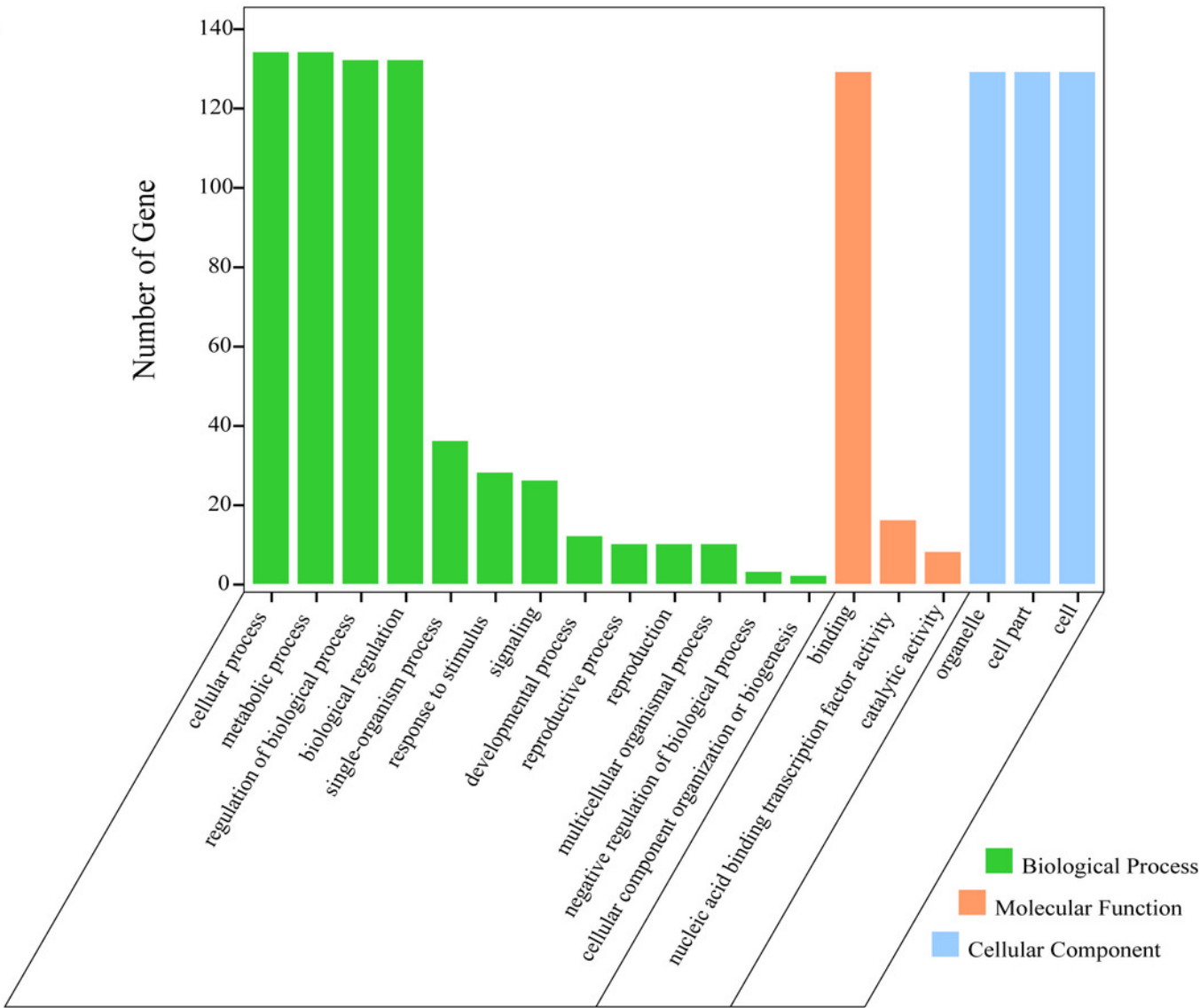



Figure 6

GO and KEGG enrichment analysis of *GhGLKs*.

(A)The numbers of level2 GO terms. (B)KEGG pathway enrichment analysis of *GhGLKs*.

A

B

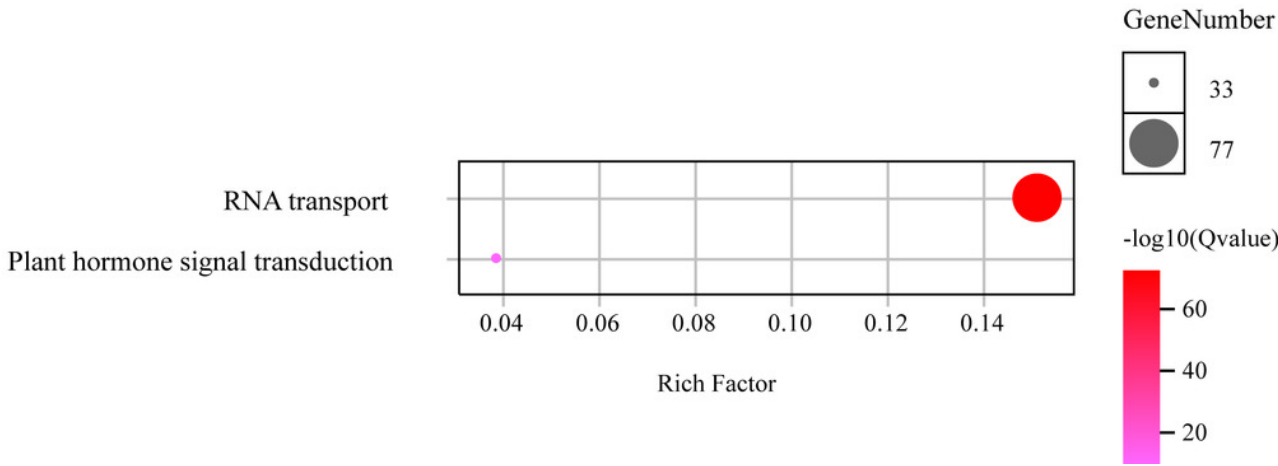



Figure 7

Differentially expressed genes of *GhGLKs* under cold, NaCl and PEG stress.

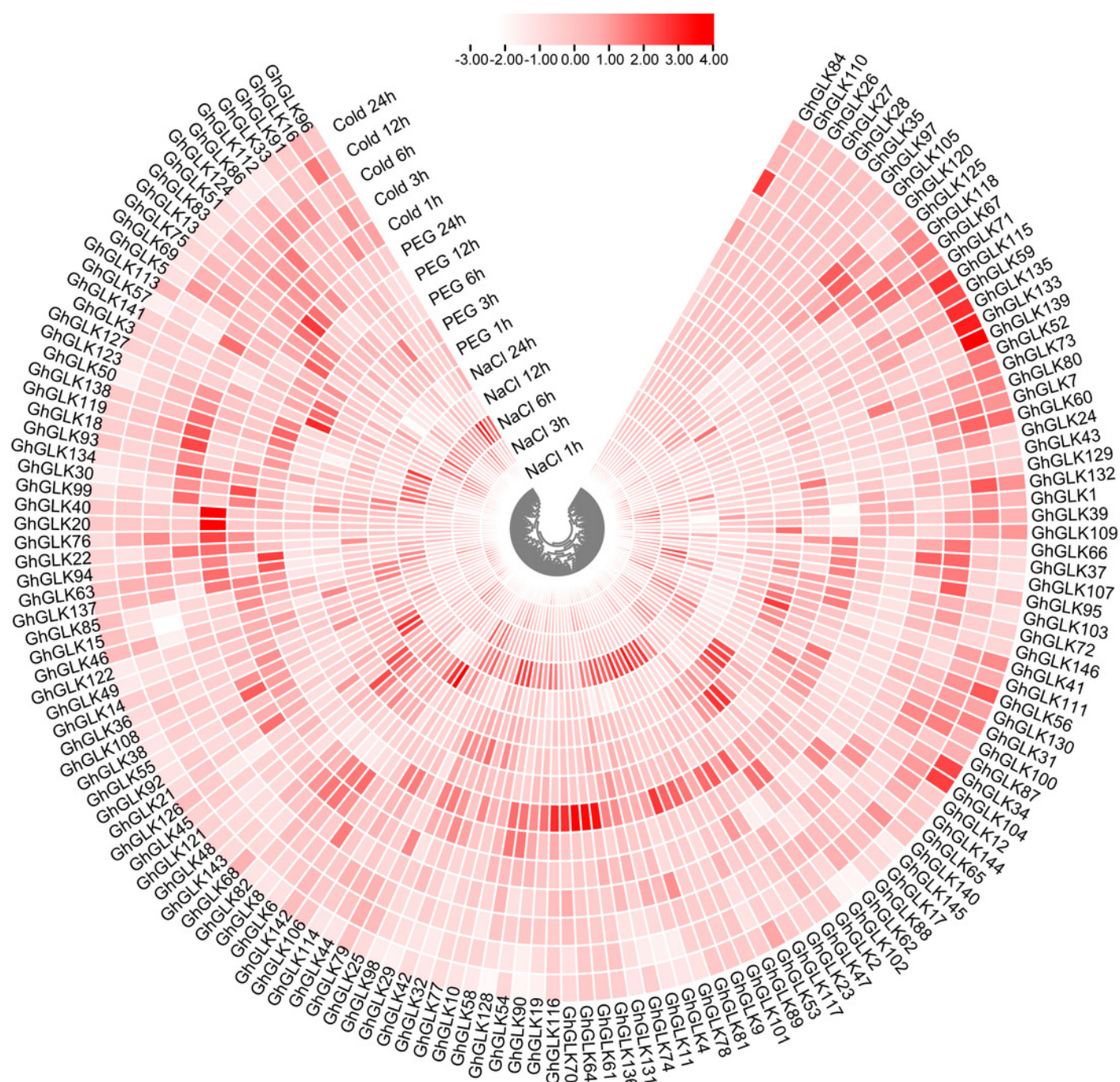



Figure 8

Expression patterns of *GhGLK* gene family members under cold, drought, and salt stress.

(A)Expression patterns analysis of *GhGLKs* under salt stress. (B)Expression patterns analysis of *GhGLKs* under drought stress. (C)Expression patterns analysis of *GhGLK*s under cold stress.