Genetics and Molecular Research

Genome-wide identification and expression analysis of *CIPK* genes in diploid cottons

J.J. Wang*, X.K. Lu*, Z.J. Yin, M. Mu, X.J. Zhao, D.L. Wang, S. Wang, W.L. Fan, L.X. Guo, W.W. Ye and S.X. Yu

State Key Laboratory of Cotton Biology,
Institute of Cotton Research of Chinese Academy of Agricultural Sciences,
Anyang, Henan Province, China

*These authors contributed equally to this study.
Corresponding author: S.X. Yu
E-mail: ysx195311@163.com

Genet. Mol. Res. 15 (4): gmr15048852
Received June 6, 2016
Accepted August 26, 2016
Published November 3, 2016
DOI http://dx.doi.org/10.4238/gmr15048852

ABSTRACT. Calcineurin B-like protein-interacting protein kinase (CIPK) plays a key regulatory role in the growth, development, and stress resistance of plants by combining with phosphatase B subunit-like protein. In the present study, *CIPK* genes were identified in the whole genomes of diploid cottons and their sequences were subjected to bioinformatic analyses. The results demonstrated that the *CIPK* gene family was unevenly distributed in two diploid cotton genomes. Forty-one *CIPK*s were identified in the D genome, mainly located on chromosomes 9 and 10, whereas thirty-nine *CIPK*s were identified in the A genome, mainly located on chromosomes 8 and 11. Based on the gene structures, *CIPK*s in cotton could be classified into two types: one that is intron-rich and the other that has few introns. Phylogenetic analysis revealed that the *CIPK* gene family members in cotton had close evolutionary relationships with those of the dicotyledonous

plants, such as *Arabidopsis thaliana* and poplar. The analysis of transcriptome sequence data demonstrated that there were differences in gene expression in different tissues, indicating that the expression of the *CIPK*s in cotton had spatio-temporal specificity. The expression analysis of *CIPK*s under abiotic stresses (drought, salt, and low temperature) in different tissues at trefoil stage demonstrated that these stresses induced the expression of *CIPK*s.

Key words: *Gossypium raimondii*; *Gossypium arboreum* L.; *CIPK* gene family; Transcriptional expression; Abiotic stress expression

INTRODUCTION

Plants have evolved many complex signaling pathways, such as the mitogen activated protein kinase (MAPK) signaling pathway and the calcium (Ca^{2+}) signal transduction pathway, to adjust to the continuous stimulation from the external environment. Research has demonstrated that Ca^{2+} signal transduction pathway participates to a great extent in many processes involved in the growth and development of plants, such as in the growth of root hairs and in guarding the movement of cells (Assmann and Wang, 2001). Moreover, plants are also affected by various abiotic stresses, such as drought, cold, salinity, hormones, and light, as well as biotic stresses, especially those induced by pathogenic bacteria (Sanders et al., 2002). Calcineurin B-like protein-interacting protein kinase (CIPK) is regulated by calcineurin B-like protein (CBL), and is one of the important components of Ca^{2+} signal transduction. The proteins encoded by CIPK genes are a family of Ca^{2+}-dependent serine/threonine kinases; these kinases have a conserved SNF kinase domain and an NAF (Asn-Ala-Phe) amino acid domain and belong to the third class of SNF1-related protein kinase 3 (SnRK3) (Harper, 2001). These two CIPK domains are required for the kinase to function. The structure of NAF is a unique domain of CIPK; it is the binding site for interaction of CBL-CIPK and plays an important role in this interaction. Extensive research has proven that the CBL-CIPK signal system plays a key role in the growth and development of plants and in their response to various stresses (Weinl and Kudla, 2009); therefore, research on the identification, structure analysis, and function of *CIPK* genes in cotton is significant in investigating the role of CBL-CIPK interaction in the growth and development of cotton plants and in the resistance of plants to stress.

Twenty-six *CIPK*s have been identified in *Arabidopsis thaliana*. There are 30, 32, 27, and 43 *CIPK*s in rice, *Sorghum*, poplar, and maize, respectively. Different members of the *CIPK* gene family respond to different stimuli in specific plant tissues and at particular developmental stages (Halfter et al., 2000). The salt overly sensitive (SOS) signaling pathway is a classical CBL-CIPK pathway involved in the signal transduction of salt stress and has been proven to be conserved, at least in poplar (Tang et al., 2010). Many other *CIPK*s are also involved in the salt-stress response; the mechanisms used by these might be different from that of the SOS pathway. A study done on *SiCIPK24* of tomato plants revealed that the main role of this gene in conferring salt stress resistance was in the transport and accumulation of more Na^+ in the plant stems (Huertas et al., 2012). Overexpression of *AtCIPK16* in *A. thaliana* and barley (*Hordeum vulgare* L.) resulted in very strong resistance of the transgenic plants to salt stress (Roy et al., 2013). Heterologous overexpression of *Brassica napus* L. *BnCIPK6* in *A. thaliana* could improve the resistance of the transgenic plants to salt stress (Chen et al.,

2012). In rice, upregulation of *OsCIPK15* expression enhanced the resistance of plants to salt stress (Xiang et al., 2007). In 2006, Xu et al. (2006) cloned and identified *AtCIPK23* from the low potassium-sensitive mutants using map-based cloning method and discovered that AtCIPK23 could directly interact with and activate the K^+ channel, AKT1, in the cell membrane. Pandey et al. (2007) found that the sensitivity of the *CIPK9* mutant to low-potassium stress was enhanced and that *AtCIPK3* might regulate the expression of the cold-resistance gene *RD29A* by regulating the transcription factor genes of CBF/DREB1 (Huang et al., 2011). Overexpression of *OsCIPK3* and *OsCIPK12* in rice significantly increased the resistance of plants to cold and drought stress. Increasing evidence demonstrates that *CIPK* genes are widely involved in the abscisic acid (ABA) signaling pathway (Chae et al., 2007), absorption of mineral nutrients, and resistance of plants to diseases and pests (Schwachtje et al., 2006).

Cotton is an important economic crop that is cultivated for oil and fiber; the cultivated species include diploid and tetraploid cotton plants, which are models for the study of plant polyploidization, cell elongation, and cell wall biosynthesis (Paterson et al., 2012). As a glycophyte, cotton displays stronger drought and salt resistance than other crops (Iqbal et al., 2011). Few studies have been conducted on the *CIPK*s in cotton and these have been mainly focused on cloning, identification, and functional analysis of a single *CIPK* (He et al., 2013). In 2012, sequencing of the "D genome" of diploid cotton *Gossypium raimondii* was completed (Wang et al., 2012) and by 2013, sequencing of the "A genome" of Asian cotton *Shixiya I* (*Gossypium arboreum* L.) was also completed (Li et al., 2014). The completion of the whole genome sequencing of diploid cotton made the first comprehensive analysis and comparison of the *CIPK* gene family members possible. Previous researchers have cloned *GhCIPK6* from *Gossypium hirsutum* L., and sequence alignment showed that this gene had homology with *AtCIPK6* of *A. thaliana.* In addition, tissue-specific expression analysis illustrated that the gene was expressed in different tissues, such as the stylus and anthers, with the expression being induced by drought, salt, and ABA (He et al., 2013). This suggests that CIPKs play a positive regulatory role in the response to salt and drought stress. It can, therefore, be concluded that CIPKs are related to the processes involved in stress, adversity, and in the growth and development of cotton.

In the present study, genome-wide identification and bioinformatic analysis of *CIPK*s was conducted to explore the homologous *CIPK*s, and to analyze their distribution in the genomes as well as their gene structure. Analyses of the expression pattern of *CIPK*s in different tissues of cotton during different growth stages were also conducted to lay a foundation for further research on the CBL-CIPK signaling network in cotton. The results of this study would be of great importance for further research on the growth and development of cotton as well as in improving their resistance to stress.

MATERIAL AND METHODS

Recognition and identification of *CIPK* family genes in cotton

To predict the members of the *CIPK* gene family in cotton, BLAST analysis was performed using the protein sequences of the *CIPK* gene families of *A. thaliana*, rice, and maize as the query sequences and the sequence data of the D and A genomes of the diploid cottons, *G. raimondii* and *G. arboreum* L., respectively; the E-value was set at 0.0001. Further confirmation was done using the screened protein sequences present in the protein database

(http://www.expasy.ch/prosite/). The protein sequences of the N-terminal SNF kinase and C-terminal NAF domains were selected as the subjects for further analysis. NAF is the specific domain of CIPK. The selection method used was as described earlier by Albrecht et al. (2001).

Phylogenetic analysis of CIPKs

The protein sequences of *A. thaliana* CIPKs were downloaded from the *A. thaliana* genome database (http://www.arabidopsis.org/), those of rice were retrieved from the rice genome database (http://rice.plantbiology.msu.edu/) and National Center of Biotechnology Information (NCBI; *Oryza sativa* subsp *japonica*), and those of maize were downloaded from the maize genome database (http://www.maizegdb.org/gbrowse/maize) and NCBI. The protein sequences of the *CIPK*s in *Populus trichocarpa* were downloaded from the genome database of *P. trichocarpa* (http://www.ncbi.nlm.nih.gov/). According to the naming conventions reported in the literature, *CIPK*s of *A. thaliana*, rice, maize, and *P. trichocarpa* were named as *AraCIPK*, *OsCIPK*, *ZmCIPK*, and *PtCIPK*, respectively. Multiple sequence alignments were performed on the sequences of the proteins encoded by all the *CIPK*s using ClustalW2. The DNAMAN software (http://www.lynnon.com/) was used for sequence alignment. The neighbor-joining method was used for the construction of the phylogenetic tree and the MEGA software (http://www.megasoftware.net/) was used for inferring the phylogenetic tree.

Prediction of the isoelectric point (pI), molecular weight, and subcellular localization of proteins encoded by *CIPK* family genes

The theoretical pI and molecular weight of CIPK were calculated using the ExpASy protein server (http://web.expasy.org/compute_pi/). The average isotopic weight of one molecule of protein was expressed in daltons (Da). The subcellular localization of each CIPK was analyzed using the CELLO v2.5 server.

Chromosome location and gene structure analysis

The sequences of *CIPK* cDNAs were obtained from the D genome of *G. raimondii* and the A genome of *G. arboreum* L., respectively. The complete D genome of *G. raimondii* and A genome of *G. arboreum* L. were compared using the cDNA sequence as a query to determine the chromosomal location of each *CIPK*; the *CIPK* cDNA and the corresponding genomic DNA sequences were compared to identify the exon/intron structure of the gene (http://gsds.cbi.pku.edu.cn/).

Transcriptome expression analysis of *CIPK*s

The expression of *CIPK*s was analyzed in three tissues, namely mature leaves, 0-DPA (day post-anthesis) ovules, and 3.0-DPA ovules of *G. raimondii* and *G. arboreum* L. The transcriptome sequencing data were obtained from the NCBI Sequence Read Archive (SRA); the registration numbers of the *G. raimondii* samples were SRX111367, SRX111365, and SRX111366, and the original sequence number of *G. arboreum* L. was SRA150181. To evaluate the expression levels of *CIPK*s, the sequence reading aligned on the sequence of *CIPK* was converted to RPKM (Mortazavi et al., 2008) and the equation used for the evaluation was as follows:

$$RPKM = 10^6 C/(NL/10^3)$$

where C refers to the reading length aligned uniquely on the transcript, N is the total reading length aligned uniquely on one specific sample, and L is the base of the transcript.

Expression analysis of *CIPKs* under different abiotic stress conditions

The seedlings of diploid cottons, *G. raimondii* and *G. arboreum* L. cv. *Shixiya 1*, were planted using the sand culture method. The true leaves, stems (hypocotyls), and roots of the seedlings were harvested and minced after abiotic stress treatments (low temperature: 4.0°C, 24 h, salt stress: 150 mM NaCl, 24 h, drought: relative water content in sand was reduced to approximately 5.0%) at the trefoil stage. Subsequently, the samples were quickly frozen in liquid nitrogen and stored at -80°C for RNA isolation. Untreated controls samples were also obtained. Total RNA was extracted from the collected tissues (Carra et al., 2007), and its concentration and purity was determined by Nanodrop2000 nucleic acid analyzer (Thermo, America). The RNA was reverse transcribed into cDNA using PrimeScript RT reagent kit with gDNA eraser (TaKaRa, China). Primer Premier 5.0 (PREMIER Biosoft) was used to design fluorescent quantitative primers (**Table S1**). Quantitative real time PCR (qRT-PCR) was performed using actin gene (GenBank accession No. AY305733) as a reference. The PCR programs were set as follows: 94°C for 30 s, followed by 40 cycles of 94°C for 5.0 s, 55°C for 34 s, and 72°C for 34 s. The assays were designed with three biological and three technical replicates. Relative quantitative analysis of the target genes was performed using the $2^{-\Delta\Delta Ct}$ method.

RESULTS

Identification of *CIPK* gene family members in the complete genome of cotton

The CBL-CIPK complex significantly affects the expression of target proteins, thereby, influencing various metabolic activities of plants. BLAST analysis was conducted using the CIPK sequences from *A. thaliana*, rice, and maize as the query sequences and the A and D genomes of cotton as the reference genomes. The results revealed that there were 39 *CIPK*s in the A genome (Table 1). On the basis of their order of presence on the chromosomes (from chromosome 1 to 13), the genes were named as *GaCIPK1*-*GaCIPK39*, with the first two letters indicating the corresponding species. There were 41 *CIPK*s in the D genome, which were similarly named as *GrCIPK1*-*GrCIPK41*. As demonstrated in Table 1, there were significant differences in the number of amino acid residues (AAs) present in the various CIPKs, which ranged from 196 to 1396. Most of the CIPKs contained between 412 and 480 AAs. GrCIPK37 was the smallest with 196 AAs because it lacked the NAF structure. The pIs ranged from 5.58 to 9.42. Except for GrCIPK2, GrCIPK15, GrCIPK22, GrCIPK30, and GaCIPK16, these five genes were slightly acidic, and the other 75 CIPKs were alkaline, similar to CIPKs of rice, *Sorghum*, and *A. thaliana*. Rice contains 30 CIPKs; except for OsCIPK1 and OsCIPK8, which were acidic, the remaining CIPKs were reported to be alkaline (Xiang et al., 2007). Of the 31 CIPKs reported in *Sorghum*, the proteins of all, except SbCIPK9, were reported to be alkaline (Li et al., 2010). The dicotyledonous *A. thaliana* contains 25 CIPKs. AtCIPK1, AtCIPK5, and AtCIPK16 were reported to be acidic whereas the others were alkaline (Li et al., 2010). The data indicate that the ancestors of CIPKs in plants were alkaline, whereas the

Table 1. Basic characteristics of CIPK genes in cotton genome.

Gene name	Chr	Accession No.	Introns	CDS (bp)	AA	pI	Mw (kDa)	Predicted subcellular localization
GrCIPK1	Chr1	Cotton_D_gene_10028535	13	1335	444	8.43	50.63	Cytoplasmic
GrCIPK2	Chr1	Cotton_D_gene_10028390	12	1128	375	5.68	43.07	Cytoplasmic
GrCIPK3	Chr1	Cotton_D_gene_10016226	0	1326	441	9.42	49.68	Cytoplasmic
GrCIPK4	Chr1	Cotton_D_gene_10015278	0	1305	434	9.28	48.79	Mitochondrial
GrCIPK5	Chr1	Cotton_D_gene_10015330	0	1347	448	8.95	50.98	Cytoplasmic
GrCIPK6	Chr3	Cotton_D_gene_10012157	0	1224	407	9.21	46.02	Mitochondrial
GrCIPK7	Chr3	Cotton_D_gene_10012160	0	1326	441	8.29	50.26	Cytoplasmic
GrCIPK8	Chr4	Cotton_D_gene_10038532	13	1338	445	6.58	49.53	Nuclear
GrCIPK9	Chr5	Cotton_D_gene_10002592	0	1293	430	6.65	49.29	Cytoplasmic
GrCIPK10	Chr5	Cotton_D_gene_10002594	0	1353	450	8.87	50.9	Cytoplasmic
GrCIPK11	Chr5	Cotton_D_gene_10005264	0	1389	462	9.19	51.2	Cytoplasmic
GrCIPK12	Chr6	Cotton_D_gene_10012417	11	1341	446	6.73	49.92	Cytoplasmic
GrCIPK13	Chr6	Cotton_D_gene_10020650	1	1374	457	8.52	52.23	Cytoplasmic
GrCIPK14	Chr6	Cotton_D_gene_10020679	13	1455	484	9.01	54.34	Cytoplasmic
GrCIPK15	Chr6	Cotton_D_gene_10015865	10	1185	394	5.85	45.19	Cytoplasmic
GrCIPK16	Chr6	Cotton_D_gene_10021141	0	1377	458	8.44	50.81	Cytoplasmic
GrCIPK17	Chr7	Cotton_D_gene_10035560	0	1308	435	8.93	49.88	Cytoplasmic
GrCIPK18	Chr7	Cotton_D_gene_10035729	13	1368	455	8.99	51.47	Cytoplasmic
GrCIPK19	Chr8	Cotton_D_gene_10015925	11	1245	414	8.91	47.28	Nuclear
GrCIPK20	Chr9	Cotton_D_gene_10037080	0	1353	450	8.95	51.19	Cytoplasmic
GrCIPK21	Chr9	Cotton_D_gene_10037140	0	1296	431	9.16	48.67	Cytoplasmic
GrCIPK22	Chr9	Cotton_D_gene_10019034	12	1131	376	5.73	42.99	Cytoplasmic
GrCIPK23	Chr9	Cotton_D_gene_10007126	0	1347	448	9.22	50.63	Cytoplasmic
GrCIPK24	Chr9	Cotton_D_gene_10033669	0	1437	478	8.54	53.36	Cytoplasmic
GrCIPK25	Chr9	Cotton_D_gene_10001863	13	1350	449	8.37	51.13	Cytoplasmic
GrCIPK26	Chr10	Cotton_D_gene_10039487	0	1296	431	9.12	48.49	Cytoplasmic
GrCIPK27	Chr10	Cotton_D_gene_10039556	0	1350	449	8.67	51.08	Cytoplasmic
GrCIPK28	Chr10	Cotton_D_gene_10040782	0	1323	440	9.31	50.06	Cytoplasmic
GrCIPK29	Chr10	Cotton_D_gene_10040781	0	1449	482	6.75	53.66	Cytoplasmic
GrCIPK30	Chr10	Cotton_D_gene_10040693	19	4191	1396	5.63	151.57	Nuclear
GrCIPK31	Chr10	Cotton_D_gene_10005466	0	1263	420	9.18	47.18	Cytoplasmic
GrCIPK32	Chr10	Cotton_D_gene_10000482	0	1389	462	8.68	52.38	Cytoplasmic
GrCIPK33	Chr11	Cotton_D_gene_10031476	0	1437	478	8.55	53.45	Nuclear
GrCIPK34	Chr11	Cotton_D_gene_10031564	13	1353	450	9.13	50.49	Cytoplasmic
GrCIPK35	Chr11	Cotton_D_gene_10036335	0	1284	427	8.46	49.03	Cytoplasmic
GrCIPK36	Chr11	Cotton_D_gene_10018680	0	1320	439	9.38	49.66	Cytoplasmic
GrCIPK37	Chr13	Cotton_D_gene_10026726	0	591	196	8.94	21.92	Cytoplasmic
GrCIPK38	Chr13	Cotton_D_gene_10022266	11	1371	456	8.53	51.21	Nuclear
GrCIPK39	scaffold84	Cotton_D_gene_10031666	0	1323	440	8.85	48.8	Cytoplasmic
GrCIPK40	scaffold131	Cotton_D_gene_10011736	13	1350	449	6.59	50.96	Cytoplasmic
GrCIPK41	scaffold163	Cotton_D_gene_10017661	12	1275	424	8.83	47.93	Cytoplasmic
GaCIPK1	CA_chr1	Cotton_A_07248	0	1239	412	9.14	46.46	Cytoplasmic
GaCIPK2	CA_chr1	Cotton_A_19207	0	1341	446	8.78	50.71	Cytoplasmic
GaCIPK3	CA_chr1	Cotton_A_13289	13	1320	439	6.86	50.27	Cytoplasmic
GaCIPK4	CA_chr4	Cotton_A_01247	1	1218	405	9.15	45.8	Cytoplasmic
GaCIPK5	CA_chr4	Cotton_A_01175	0	1353	450	8.95	51.25	Cytoplasmic
GaCIPK6	CA_chr4	Cotton_A_32557	0	1377	458	8.44	50.78	Cytoplasmic
GaCIPK7	CA_chr4	Cotton_A_15642	13	1155	384	6.35	43.67	Cytoplasmic
GaCIPK8	CA_chr5	Cotton_A_04333	1	1554	518	8.38	58.75	Cytoplasmic
GaCIPK9	CA_chr5	Cotton_A_04331	0	1293	430	6.45	49.2	Cytoplasmic
GaCIPK10	CA_chr5	Cotton_A_08366	0	1377	458	9.03	50.77	Cytoplasmic
GaCIPK11	CA_chr6	Cotton_A_37254	0	1431	476	8.26	53.08	Cytoplasmic
GaCIPK12	CA_chr6	Cotton_A_25899	10	1098	365	9	41.59	Cytoplasmic
GaCIPK13	CA_chr7	Cotton_A_05019	0	1389	462	8.68	52.42	Cytoplasmic
GaCIPK14	CA_chr7	Cotton_A_04154	0	1224	407	9.21	46.07	Mitochondrial
GaCIPK15	CA_chr7	Cotton_A_04158	0	1326	441	8.63	50.41	Cytoplasmic
GaCIPK16	CA_chr7	Cotton_A_25709	19	4179	1392	5.58	151.06	Nuclear
GaCIPK17	CA_chr8	Cotton_A_17423	13	1335	445	8.98	50.43	Mitochondrial
GaCIPK18	CA_chr8	Cotton_A_38153	0	1440	479	7.09	53.45	Cytoplasmic

Continued on next page

acidic CIPKs evolved later. The molecular weight of CIPKs ranged from 21.92 to 151.57 kDa, with the majority of proteins being in the range from 41.59 to 58.75 kDa. GrCIPK37 had the minimum molecular weight of 21.92 kDa. The molecular weights of GaCIPK36, GaCIPK16, and GrCIPK30 were 81.96, 151.06, and 151.57 kDa, respectively. Two CIPKs, on each of the D (GrCIPK4 and GrCIPK6) and A (GaCIPK14 and GaCIPK17) genomes, were predicted to be localized to the mitochondria. In contrast, the proteins of five *CIPK*s on the D genome (GrCIPK8, GrCIPK19, GrCIPK30, GrCIPK33, and GrCIPK38) and four CIPKs on the A genome (GaCIPK16, GaCIPK21, GaCIPK28 and GaCIPK36) were predicted to localize to the nucleus. The remaining CIPKs were localized in the cytoplasm, suggesting that the subcellular localization of the proteins of *CIPK* gene family members in cotton was diverse. It is, therefore, suggested that their corresponding functions should also have diversity, and hence, their subcellular location needs further experimental verification. Previous studies reported that PsCIPK was localized in cytosol and outer membrane in leguminous plants, which was confirmed by immunofluorescence and confocal microscopy (Mahajan et al., 2002). AtCIPK1 could be located in the cell membrane, cytoplasm, and nucleus.

Functional domain analysis of the CIPK family members in cotton

Functional domain analysis of the identified CIPK family members was conducted (http://prosite.expasy.org/) and the results revealed that all the protein sequences of CIPKs in the D and A genomes contained the protein kinase domain (PS50011). Except for GrCIPK37 in the D genome, all CIPKs contained the NAF domain (PS50816). Although GrCIPK37 had no NAF domain, Wang et al. (2012) annotated it as a CIPK; therefore, we also listed it as a CIPK family member in the present study and analyzed it with the other members. Proteins of 41 CIPKs in the D genome and 37 CIPKs (except for GaCIPK6 and GaCIPK12) in the A genome contained the serine/threonine protein kinase active site (PS00108). Except for GrCIPK2, GrCIPK7, and GrCIPK22 in the D genome and GaCIPK7, GaCIPK12, and GaCIPK15 in the A genome, all CIPKs encoded proteins with a protein kinase ATP binding domain (PROTEIN_KINASE_ATP, PS00107). All the CIPKs in the D and A genomes, except for GaCIPK12, contained a proton acceptance locus (ACT_SITE). There were some differences in the distribution of CIPKs and a number of CIPK family members contained five domains; it is speculated that these five domains have different roles in CIPK function. In addition, protein encoded by *GrCIPK12* in the D genome and GaCIPK34 in the A genome contained the ribosomal protein L29 marker (RIBOSOMAL_L29, PS00579) and GrCIPK24 contained an EF-hand type calcium-binding domain (EF_HAND_1, PS00018). In contrast, the zinc finger ring type domain (ZF_RING_2, PS50089) appeared in GaCIPK36 but was not found in other CIPK family members. It is speculated that these genes experienced directional evolution in structure resulting from changes in the environment. These structures might have special functional properties, which would need further experimental verification.

Distribution of *CIPK* gene family members in the complete genome of cotton

The information about the position of genes on chromosomes provides important evidence for the study of evolution and the function of a gene family. Combined with the chromosome information of the cotton A and D genomes and the locations of the *CIPK*s on the chromosomes, the distribution map of *CIPK*s on the chromosomes was prepared (Figure 1).

Table 1. Continued.

Gene name	Chr	Accession No.	Introns	CDS (bp)	AA	pI	Mw (kDa)	Predicted subcellular localization
GaCIPK19	CA_chr8	Cotton_A_39763	0	1323	440	9.35	50.06	Cytoplasmic
GaCIPK20	CA_chr8	Cotton_A_15608	0	1308	435	9.26	48.91	Cytoplasmic
GaCIPK21	CA_chr8	Cotton_A_15585	11	1368	456	8.36	51.31	Nuclear
GaCIPK22	CA_chr8	Cotton_A_32831	0	1296	431	9.12	48.55	Cytoplasmic
GaCIPK23	CA_chr8	Cotton_A_36065	14	1371	456	9.09	51.94	Cytoplasmic
GaCIPK24	CA_chr8	Cotton_A_40954	0	1314	437	8.58	49.37	Cytoplasmic
GaCIPK25	CA_chr9	Cotton_A_32356	0	1275	424	8.91	48.66	Cytoplasmic
GaCIPK26	CA_chr9	Cotton_A_02928	13	1302	434	8.98	48.73	Cytoplasmic
GaCIPK27	CA_chr9	Cotton_A_23852	0	1320	439	9.39	49.62	Cytoplasmic
GaCIPK28	CA_chr9	Cotton_A_17529	0	1437	478	8.2	53.43	Nuclear
GaCIPK29	CA_chr10	Cotton_A_11022	13	1323	440	6.49	50.26	Cytoplasmic
GaCIPK30	CA_chr10	Cotton_A_19508	13	1350	449	8.56	51.23	Cytoplasmic
GaCIPK31	CA_chr10	Cotton_A_04065	0	1326	441	9.41	49.73	Cytoplasmic
GaCIPK32	CA_chr10	Cotton_A_17083	0	1347	448	9.27	50.73	Cytoplasmic
GaCIPK33	CA_chr11	Cotton_A_12475	0	1308	435	9	49.78	Cytoplasmic
GaCIPK34	CA_chr11	Cotton_A_16517	11	1317	439	6.15	49.15	Cytoplasmic
GaCIPK35	CA_chr11	Cotton_A_34885	13	1350	449	7.15	51.01	Cytoplasmic
GaCIPK36	CA_chr11	Cotton_A_23687	16	2193	730	8.45	81.96	Nuclear
GaCIPK37	CA_chr11	Cotton_A_01750	13	1329	442	6.91	50.66	Cytoplasmic
GaCIPK38	CA_chr12	Cotton_A_30062	1	1374	457	8.52	52.37	Cytoplasmic
GaCIPK39	CA_chr13	Cotton_A_09063	0	1296	431	9.14	48.63	Cytoplasmic

AA = amino acid amount; PI = isoelectric point; Mw = molecular weight, kilodalton (kDa) is used as unit.

As shown in Figure 1A, 38 of the 41 *CIPK*s were mapped to 11 chromosomes of the D genome of cotton. In comparison, distribution of the remaining three *CIPK*s was as follows: *GrCIPK39*, *GrCIPK40*, and *GrCIPK41* were located on scaffold84, scaffold131, and scaffold163, respectively. None of the three genes was located on the corresponding chromosomes. Chromosomes 2 and 12 had no *CIPK*s; conversely, chromosome 10 had the highest number of *CIPK*s (seven), followed by chromosome 9 with six *CIPK*s. Chromosomes 1 and 6 had five *CIPK*s and chromosomes 4 and 8 had one. As shown in Figure 1B, all *CIPK*s in the A genome of cotton were located on the chromosomes and were found on the 11 chromosomes; however, neither chromosome 2 nor chromosome 3 contained any *CIPK*. Chromosome 8 contained the maximum number (eight) of *CIPK*s, followed by chromosome 11 with five genes. In contrast, chromosomes 12 and 13 contained only one *CIPK*. The remaining chromosomes contained 2-4 *CIPK*s each. In the D genome, *GrCIPK6* and *GrCIPK7* were distributed on chromosome 3 in the form of a cluster. *GrCIPK9* and *GrCIPK10* were located on chromosome 5 in a close-linking mode. These two gene clusters contained two gene pairs, *GrCIPK6*/*GrCIPK9* and *GrCIPK7*/*GrCIPK10*, which had fewer intron types and similar functions. Furthermore, *GaCIPK8* and *GaCIPK9* were closely linked on chromosome 5 on the A genome, and *GaCIPK15* and *GaCIPK14* were distributed on chromosome 7 in a cluster, both of which contained two gene pairs, *GaCIPK8*/*GaCIPK15* and *GaCIPK9*/*GaCIPK14*, with fewer intron types and similar functions. This distribution pattern might have been caused by the substitution and insertion in chromosomes. A gene family is created by the random amplification of genes, resulting in the formation of gene clusters. The scattered distribution of the members of a gene family on many chromosomes is most likely caused by partial-fragment replication of chromosome regions (Schauser et al., 2005). Compared to other eukaryotes, plants have a higher gene duplication rate (Wei et al., 2014). The results of previous studies showed that gene duplication and separation of the latter stages are the two main objectives of evolution (Chothia et al., 2003), resulting in the diversity of gene family members. Wang et al. (2012) demonstrated that in *G. raimondii*, complete genome replications occurred at least

two times. The scattered distribution mode of the *CIPK* genes in the complete chromosome might reflect a series of complete genomes, chromosomes, and large-fragment duplication events with typical characteristics of the D and A chromosome genomes. Duplication of genes leads to diversity of gene functions, which plays a very important role driving the evolution of new features, organ differentiation, and better adaptation to changes in environments (Flagel and Wendel, 2009). Phylogenetic analysis reveals that *G. arboreum* L. and *G. raimondii* were derived from a common ancestor dating about 5.0 mya, that the genomes of these two species exhibited a high degree of collinearity at the chromosome level, and that the number of genes and sequences were very similar (Li et al., 2014). For example, there were 13 *CIPK*s on chromosomes 9 and 10 of the D genome, and 13 *CIPK*s on chromosomes 8 and 11 of the A genome, which might also be the result of gene duplication or partial segment duplication over the long evolutionary history of the cotton genome.

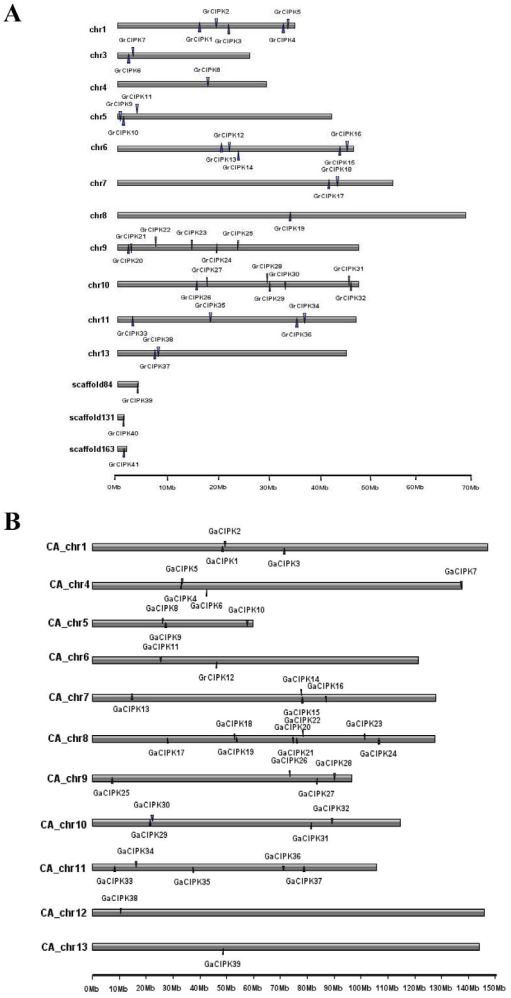

Figure 1. Chromosome distribution of the *CIPK* genes in the cotton genome. **A.** *CIPK* genes in *Gossypium raimondii*; **B.** *CIPK* genes in *Gossypium arboreum* L.

Genetic structural analysis of the *CIPK* gene family members in cotton

Gene structure analysis is important in the research on gene evolution. According to the number of introns, the *CIPK* gene family members of the D and A genomes of cotton were divided into two categories (as shown in Figure 2), one with more than 10 introns and the other with fewer than three introns. In contrast, the category containing less than three introns had 26 members in the D genome, accounting for 63.41% of the members. Except for *GrCIPK13*, which contained two exons, all the 25 members contained only one exon. The category with more than 10 introns comprised 15 members in the D genome. *GrCIPK15* contained only 11 exons and *GrCIPK30* contained 20 exons. The rest of the *CIPK* members had between 12 and 14 exons each. The situation in the A genome was similar to that in the D genome. The category with less than three introns comprised 25 members, accounting for 64.10% of the members, of which *GaCIPK4*, *GaCIPK8*, and *GaCIPK38* contained two exons and the remaining 22 members contained one exon. The category with more than 10 introns contained 14 members and more than 10 exons. *GaCIPK12* had only 11 exons. In contrast, *GaCIPK16* had the highest number (20) of exons. It is speculated that the number of exons of each *CIPK* member in the D and A genomes of cotton was similar during the evolutionary process of *CIPK* gene structure, which was relatively conserved, suggesting that the functions of these genes might also be consistent. The results of previous studies showed that the insertion of small fragments of DNA could change the function of the gene, and that the gene could disappear by natural selection (Long et al., 1995). *CIPK* members, *GrCIPK30* and *GaCIPK16*, of the cotton D and A genomes, respectively, had the highest number (up to 20) of exons, with large differences in the exon length, suggesting that there were significant changes in the structures or functions of these two genes in the evolutionary process of cotton. In addition, this indicates that there were relatively stable exon-intron pairs in either the D or A genome during the evolutionary process of cotton. Compared to the introns, the exons were more vulnerable to selective pressures from the external environment. The structures of the exons in many replication-type gene families were generally conserved; therefore, the differences in the structures of exons and introns caused by insertion-deletion events could be used to predict the evolutionary history of gene families (Lecharny et al., 2003).

Gene pair analysis of the *CIPK* gene family

The results of cluster analysis conducted on the D genome revealed that 41 members of *CIPK* in the D genome could be divided into 2 categories, one with no or very few introns and the other with several introns (Figure 3A). Similar results were also obtained with the A genome (Figure 3B). In addition, there were 13 pairs of homologous genes in the D and A genomes (**Table S2**).

The cluster analysis of the D and A genomes (Figure 3C) also illustrated that 36 *CIPK*s in the D genome were paired with 36 *CIPK*s the A genome (**Table S2**). The formation of such gene pairs might have been caused by duplication events. The sequence consistency of proteins encoded by the CIPK gene pairs of the D genome was from 65.53% (*GrCIPK7*/*GrCIPK10*) to 92.2% (*GrCIPK25*/*GrCIPK4*0), with an average of 82.51% consensus in the sequences. By comparison, the proteins encoded by the *CIPK* gene pairs of the A genome had a sequence consistency from 65.53% (*GaCIPK8*/*GaCIPK15*) to 92.71% (*GaCIPK3*/*GaCIPK29*), with an average consensus of 82.79%.

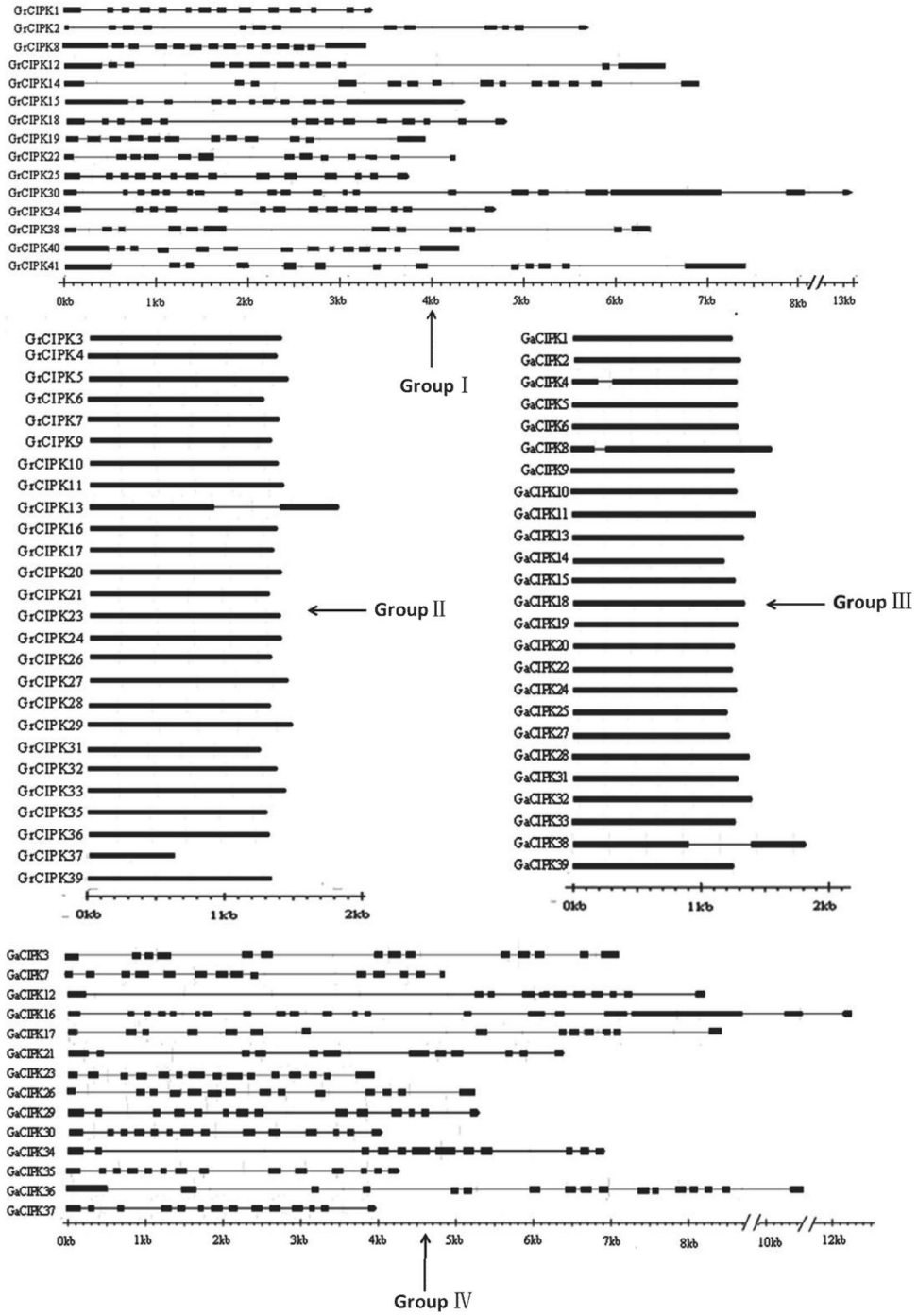

Figure 2. Intron-exon structure analysis of the *CIPK* gene family in cotton. Group I and Group II, *CIPK* genes in *Gossypium raimondii*; Group III and Group IV, *CIPK* genes in *Gossypium arboreum* L.

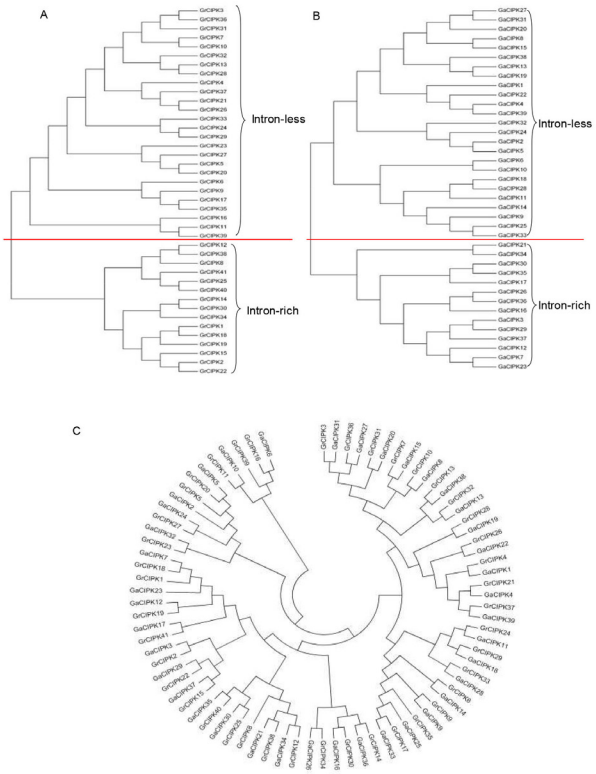

Figure 3. Phylogenetic analysis of the CIPK proteins in the D and A genomes in cotton. **A.** Phylogenetic trees of the CIPK proteins in D genome. **B.** Phylogenetic trees of the CIPK proteins in A genome. **C.** Phylogenetic trees of the CIPK proteins in D and A genomes.

In contrast, the consensus in the sequences of corresponding proteins encoded by *CIPK* family gene pairs of the D and A genomes ranged from 89.05% (*GrCIPK14*/*GaCIPK36*) to 99.73% (*GrCIPK2*/*GaCIPK3*), with an average consensus of 97.47%. Of these pairs, seven showed >99.00% consensus; these pairs were as follows: *GrCIPK32*/*GaCIPK13*, *GrCIPK26*/*GaCIPK22*, *GrCIPK37*/*GaCIPK39*, *GrCIPK6*/*GaCIPK14*, *GrCIPK34*/*GaCIPK26*, *GrCIPK40*/*GaCIPK35*, and *GrCIPK2*/*GaCIPK*3. These results indicate that the interspecies genetic relationship between the *CIPK* gene family members of cotton is stronger than the intraspecies relationship. They also suggest that the common ancestor of *G. arboreum* L. and *G. raimondii* formed the *CIPK* gene family during the evolutionary process. After the differentiation of the two species about 5.0 mya from the same progenitor, the *CIPK* gene family experienced linear synchronous evolution in the two species. These comprehensive analyses demonstrate that a gene duplication event had a significant effect on the amplification of the *CIPK* gene family members in cotton, and that the amplification of the *CIPK* members might be the result of a high-stress environment. Cotton is chronically affected by an adverse environment during the growth period, which promotes the widespread amplification of genes with similar functions. Moreover, the newly formed CIPK genes must have evolved with the corresponding *CBL* genes. These functional limitations might have helped in the evolution of the current CBL-CIPK signaling network (Kolukisaoglu et al., 2004).

Phylogenetic analysis of the *CIPK* gene families of cotton and four other species

To test the evolutionary relationship between the *CIPK* gene family members in cotton and those in *Arabidopsis*, rice, maize, and poplar, the amino acid sequences of all CIPK members from these species were compared and a phylogenetic tree was constructed (see Material and Methods). The analysis of the phylogenetic tree revealed that CIPK of the D and A genomes of cotton were mostly clustered together (Figure 4). The relationship between *G. raimondii* and *G. arboreum* L. was closer than between the other four species. In addition, many members of the *CIPK* gene family of cotton and poplar clustered together, suggesting that these two species were evolutionarily close; this was followed by the evolutionary closeness of cotton with *Arabidopsis* (**Table S3**). These genes are similar in structure and it is predicted that their functions might also be similar. There were more family members in cotton than in *Arabidopsis*, rice, and poplar plants, suggesting a specific linear amplification of the gene family in cotton. Whether these additional members of the gene family have additional functions, as well, or whether they are only produced because functional redundancy requires further experimental verification.

Figure 4. Phylogenetic analysis of the CIPK proteins in cotton, *Arabidopsis*, rice, maize, and poplar. The species used for the construction of the phylogenetic tree are Gr, *Gossypium raimondii*; Ga, *Gossypium arboreum* L.; At, *Arabidopsis*; Os, rice; Zm, maize; and Pt, poplar.

CIPK expression analysis based on the transcriptome sequencing

Transcription analysis has been used for identifying the protein-coding genes during the annotation of the D and A genomes (Wang et al., 2012; Li et al., 2014). In this study, the sequence read lengths of the matched *CIPK* genes were converted to RPKM values and used to predict the gene expression levels. The data were retrieved from NCBI and at least 20 nucleotide sequences were selected and used in this study.

The abundance of transcripts in the cotton leaves at the beginning of flowering stage, and in 0 and 3.0 DPA ovules was detected (Figure 5). The results showed that about 9.76% of the *CIPK* genes (*GrCIPK12*, *GrCIPK18*, *GrCIPK30*, and *GrCIPK31*) in the D genome were expressed in the leaves (Figure 5A). Of these, the expression of *GrCIPK18* and *GrCIPK31* in the leaves was much higher than that of *GrCIPK12* and *GrCIPK30*, indicating that these two genes play an important role in leaf development and that *GrCIPK18* and *PtCIPK13* are homologous. Nine *CIPK* genes (21.95%) were expressed in the 0-DPA ovules and eleven (26.83%) were expressed in the 3.0-DPA ovules. Compared to the expression in the 0-DPA ovules, the expression of *GrCIPK4*, *GrCIPK7*, *GrCIPK8*, *GrCIPK30*, *GrCIPK31*, *GrCIPK36*, and *GrCIPK40* in the 3.0-DPA ovules was increased by 1.19-, 1.09-, 1.02-, 3.02-, 3.87-, 2.82-, and 1.78-times, respectively, indicating that these genes play a major role in fiber elongation. *GrCIPK12*, *GrCIPK18*, *GrCIPK23*, and *GrCIPK26* were all downregulated with different extent in 3.0-DPA ovules compared with that in 0-DPA ovules, indicating that these genes play important roles in the initiation of fiber formation in cotton. Of the *CIPK* genes in the A genome, approximately 41.03% (sixteen) were expressed in the leaves, with five genes being highly expressed in the leaves of *G. arboreum* L. (Figure 5B). *GaCIPK8*, *GaCIPK14*, *GaCIPK22*, *GaCIPK33*, *GaCIPK39*, *GaCIPK8*, and *OsCIPK2* were observed to be homologous. In contrast, 35.90% (fourteen) of the CIPK genes were expressed in the 0-DPA ovules, with five genes, *GaCIPK4*, *GaCIPK8*, *GaCIPK14*, *GaCIPK33*, and *GaCIPK39*, being highly expressed. One-third (13) of the *CIPK* genes were expressed in the 3.0-DPA ovules in *G. arboreum* L., and two of these, namely *GaCIPK8* and *GaCIPK39*, were highly expressed. Compared to the expression in 0-DPA ovules, only *GaCIPK2* was significantly upregulated (1.02-times) in the 3-DPA ovules, indicating that this gene plays a main role in fiber elongation. Compared to the expression in the 0-DPA ovules, the expression of three *CIPK* genes, namely *GaCIPK7*, *GaCIPK14*, and *GaCIPK32*, was downregulated in the 3.0-DPA ovules by 1.95-, 1.54-, and 1.27-times, respectively, indicating that these three genes play important roles in the initiation of cotton fiber. The expression of another nine *CIPK* genes (*GaCIPK4*, *GaCIPK8*, *GaCIPK10*, *GaCIPK15*, *GaCIPK20*, *GaCIPK28*, *GaCIPK31*, *GaCIPK33*, and *GaCIPK39*) in the 0- and 3.0-DPA ovules was similar, suggesting that they all play important roles in the initiation and development of fiber. *GaCIPK22* and *GaCIPK25* were expressed only in the leaves, and *GaCIPK4* was expressed only in the 0-DPAovules, indicating that the expression of these three genes was tissue specific. *GaCIPK22* and *GrCIPK26* belong to homologous gene pairs, and *GaCIPK22* was abundantly expressed only in the leaves. The expression of *GrCIPK26* in the 0-DPA ovules was high, suggesting that homologous genes might also have functional diversity.

Expression analysis of the *CIPK* genes in diploid cottons under different stress conditions

To investigate the anti-adversity function of *CIPK* genes in cotton, their expression

patterns were studied in different tissues of diploid cottons exposed to different stress conditions (low temperature, drought, and salt stress). The results showed that 80 *CIPK* genes were expressed in the roots, stems, and leaves of diploid cottons, but the expression levels of most of the genes were different in the roots, stems, and leaves after exposure to abiotic stresses (Figure 6).

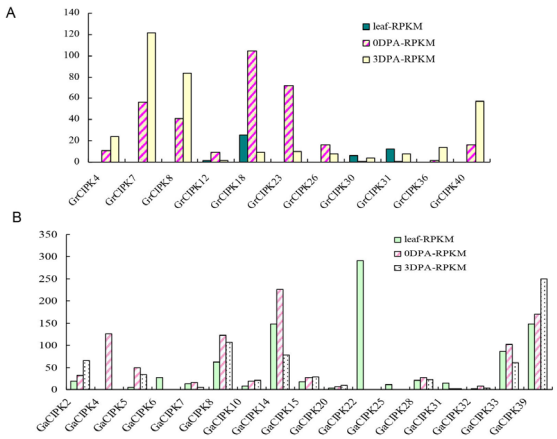

Figure 5. Transcript abundance analysis of the *CIPK* genes in mature leaves, 0-DPA and 3.0-DPA ovules in *Gossypium raimondii* and *Gossypium arboreum* L. The transcript expressions were calculated using the RPKM method (Mortazavi et al., 2008).

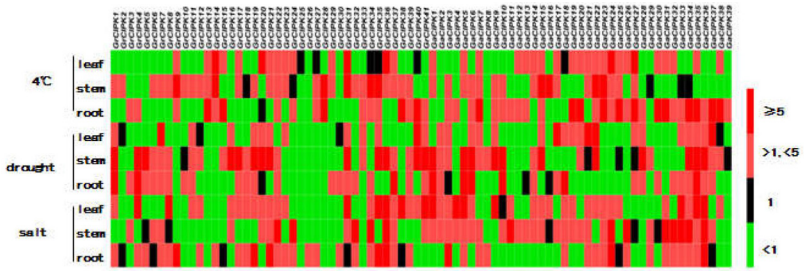

Figure 6. Expression of the *CIPK* genes in different tissues of diploid cotton under different abiotic stress conditions. Gene expression analysis of three different tissues of diploid cotton species was launched under different stresses with qRT-PCR method. Red color represented genes upregulated and the expression folds were greater than or equal to 5; Pale red color represented the genes upregulated with expression folds was greater than 1 and less than 5; black color represented the genes with expression amount was almost unchanged; green color represents genes downregulated and the expression folds were less than 1.

Sixteen, 28, and 14 *CIPK* genes were upregulated in the roots, stems, and leaves of *G. raimondii*, respectively and 3, 11, and 22 genes were downregulated, respectively, after 4°C treatment. In the roots, stems, and leaves of *G. arboreum* L. 29, 15, and 22 *CIPK* genes were upregulated and 10, 19, and 16 genes were downregulated, respectively, after 4°C treatment. However, the expression levels of *CIPK* genes varied. Moreover, similar expression patterns of *CIPK* genes were observed under both drought and salt stress. In different tissues, the number of *CIPK* genes up- or downregulated was different, showing tissue specificity. The results also showed that 11 *CIPK* genes in *G. raimondii* and *G. arboreum* L. were upregulated

in the leaves, stems, and roots simultaneously after low temperature treatment (**Table S4**). In addition, 12 and 15 *CIPK* genes were simultaneously upregulated in the leaves, stems, and roots in both the diploid cotton species after drought and salt treatment, respectively. These results suggest that these genes could be induced by low temperature, drought, and salinity, and thus, play an important role in the resistance of cotton to the imposed stress. Moreover, differences in the expression of *CIPK* genes under different stress conditions reveal specificity in the expression.

Analysis of anti-adversity function of *CIPK* genes in cotton showed that 9 (*GrCIPK20/21/23/31*, *GaCIPK15/19/21/22/35*), 14 (*GrCIPK6/7/21/22/31/36/37*, *GaCIPK3/5/6/7/10/14/15*), and 7 (*GrCIPK22/34/39 GaCIPK32/33/34/36*) genes showed upregulation in leaves, stems, and roots, respectively, after low temperature, drought, and salt stress, indicating that these genes could respond to various adversities. Therefore, we speculate that these *CIPK* genes could respond to many other stresses, and therefore, play a vital role in stress resistance.

DISCUSSION

In recent years, with the development of genome research, comparative genomics has been extensively used for research on gene families, which have attracted increased attention in many species. Several gene families, such as the *TIFY* (He et al., 2015) and *MAPKKK* (Yin et al., 2013) gene families, have been identified in cotton. As a pioneer crop in saline soils, the adaptability of cotton is widespread. *CIPK* genes are important for stress resistance. The previous transcriptome sequencing experiments demonstrated that some *CIPK* genes of cotton were associated with resistance to cold, salt, and high temperatures. Considering the fact that little research has been conducted on the *CIPK* gene family in cotton, in the present study, we identified *CIPK* genes present in the complete genomes (the D and A genomes) of diploid cotton. The *CIPK* gene members of cotton were assessed for their gene structure, location, evolutionary relationships, and transcription expression patterns, and their response to different abiotic stresses was also evaluated. The results obtained should lay the foundation for an in-depth study of the molecular functions of *CIPK* genes during the later stages of cotton growth.

In the present study, bioinformatic analyses of the D and A genomes of diploid cotton were conducted using the protein sequences of the *CIPK* gene family of *Arabidopsis*, rice, and corn as the query sequences. Forty-one gene members were identified in the D genome and 39 members were identified in the A genome of cotton, which are relatively more evolutionarily conserved and have more members in the gene family than in *Arabidopsis*, rice, and poplar. Gene structure analysis showed that the *CIPK* gene family of cotton could be divided into two types: one that is intron-rich and the other that is intron-less or has few introns. The intron-rich type contains more than 10 exons, and the no-intron type (intron-less) contains 1-2 exons. Moreover, the CIPK gene family members of cotton also had complex subcellular localization patterns, which might be due to the directional evolution of function and structure of the *CIPK* genes over long-term. Conserved domain analysis revealed that CIPK members retained relatively conserved domains, and all the sequences of CIPK in the D and A genomes contained a protein kinase domain (PS50011). Most CIPKs contained an NAF domain (PS50816), the active site (PS00108) of serine/threonine protein kinases, the ATP-binding region of protein kinases (PROTEIN_KINASE_ATP, PS00107), and a proton acceptance locus (ACT_SITE). These domains might be an important prerequisite for the functions of the CIPK and might also be a factor in changing the structures and functions of *CIPK* genes. The participation of

various structural domains in metabolic regulation and their specific functions need further experimental verification. The functions of two genes, GrCIPK24 and GaCIPK36, in response to abiotic stress might be due to their special domains, namely EF-type calcium-binding domain (EF_HAND_1, PS00018) of GrCIPK24 and zinc ring-type domain (ZF_RING_2, PS50089) of GaCIPK36. However, further research is needed to confirm this observation.

Although the evolution of one gene family can be regulated by several mechanisms, evolutionary and structural analyses can help in determining the origin and relationships of different species. In this study, the phylogenetic tree analysis showed that the *CIPK* gene family members of the D and A genomes of cotton were evolutionarily close. The *CIPK* family members of the dicotyledonous *Arabidopsis* and poplar were also closely related, unlike the members of the monocotyledonous maize and rice. *AtCIPK3* and *GrCIPK2*/*GaCIPK3*, *AtCIPK6* and *GrCIPK26*/*GaCIPK22*, *AtCIPK21* and *GrCIPK8*, and *AtCIPK23* and *GrCIPK30*/*GaCIPK16* were clustered together and formed homologous gene pairs. The clustered genes that formed homologous gene pairs between cotton and rice included *OsCIPK2* and *GrCIPK10*/*GaCIPK8*, and *OsCIPK5* and *GrCIPK13*/*GaCIPK38*; the homologous gene pairs in the case of cotton and maize were *ZmCIPK5* and *GrCIPK13*/*GaCIPK38*, *ZmCIPK15* and *GrCIPK32*/*GaCIPK13*, and *ZmCIPK23* and *GrCIPK14*/*GaCIPK36*. The results indicate that these genes might have similar functions. Previous studies have revealed that the AtCIPK3 might regulate the expression of the cold resistance gene, RD29A, by regulating the gene for the transcription factor, CBF/DREB1 (Huang et al., 2011). In addition, *GrCIPK2* was only upregulated in the shoots after low temperature and was downregulated in response to other stresses (drought and salt stress). *GaCIPK3* was upregulated in all the tissues under salt, drought, and low-temperature treatments except in the leaves after low-temperature treatment, where it was downregulated. Although *GrCIPK2* and *GaCIPK3* could form a gene pair as suggested by their structures, the function of these two genes started to differ, possibly due to stronger resistance and adaptability of Asian cotton (one of the four cultivated species) in the process of long-term artificial selection. *AtCIPK6* is involved in the growth and development of plants (Tripathi et al., 2009) and is regulated under different adverse conditions (Chen et al., 2013). *GrCIPK26* was up-regulated (by 2.09-times) in the root after salt stress treatment; *GaCIPK22* was upregulated in the roots, stems, and leaves after low temperature treatment, and was upregulated in the leaves and stems after drought treatments. The *GrCIPK26/GaCIPK22* pair showed similar function as the homologous *AtCIPK6* and could be induced under a variety of stress conditions. Gene pair *GrCIPK14/GaCIPK36* showed a similar function in response to salt stress and had a significantly different response to drought stress, indicating that the anti-adversity function of these two genes had already begun to differentiate. Furthermore, the expression of *OsCIPK2* and *OsCIPK5* could be induced by drought stress and ABA (Xiang et al., 2007). The expression of *ZmCIPK5*, *ZmCIPK15*, and *ZmCIPK23* was upregulated under high- and low-temperature stress (Chen et al., 2011).

These results indicate that many members of the *CIPK* gene family in cotton play important roles in the growth and development of plants and in their reaction to stress, but the specific mechanism by which they perform these roles needs further elucidation. Cotton fibers originate from the epidermal cells of ovules, and their growth and development are divided into four stages, namely initiation, elongation, secondary cell wall synthesis, and a mature period. The fiber cell differentiation directly affects the length of the mature fibers on the ovules, and the initial stage of fiber development is from 2.0 to 1.0 DPA. The elongation stage is from 2.0 to 8.0 DPA (Wang et al., 2010) and 0-3.0 DPA is the critical period for the

formation of fiber (lint) from the differentiated epidermal cells on the ovules. We, therefore, selected 0 DPA as the initial stage and 3.0 DPA as the elongation stage of fiber development for our analyses. The expression analysis of *CIPK* based on transcriptome sequencing data demonstrated that there were differences in the *CIPK* transcript levels in the leaves during the initial flowering period and in the ovules at different times of flowering. The differences in the expression were observed for the different *CIPK* gene family members in the D or A genome, suggesting that these genes had temporal and spatial specificity of expression during the growth and development processes in cotton. However, there were more *CIPK* members in the A genome whose transcript expression in the ovule was altered, indicating that the role of *CIPK* genes in the A genome is more important in the development of fiber than those of the genes in the D genome. This could be due to the fact that *G. arboreum* L. has more well-developed fibrous tissue than *G. raimondii* and has more genes involved during the developmental process of fiber. This study showed that *GrCIPK8* and *GrCIPK30* play a role in cotton fiber elongation, and that *GrCIPK8*/*AtCIPK21*, and *GrCIPK30* and *AtCIPK23* are homologous genes. It is reported that *AtCIPK23* could regulate potassium uptake and stomatal movement (Xu et al., 2006). Thus, we speculate that the functions of *GrCIPK8* are diverse. *GrCIPK26* and *AtCIPK6* are homologous genes and *GrCIPK26* plays a role in the initiation of cotton fiber. *AtCIPK6* responds to abiotic stress and is regulated by ABA (Chen et al., 2013), suggesting that *GrCIPK26* also has a variety of functions in the growth and development of cotton. The current research demonstrates that the CBL-CIPK system plays an important role in the mechanism of plant resistance to stress; however, research on the CBL-CIPK signal network has mostly been focused on the model plants, like *Arabidopsis* and rice. Whether the CBL-CIPK signaling system of *Arabidopsis* and rice exists in other monocotyledonous and dicotyledonous plants, and whether there are differences in the expression and functions of *CBL* and *CIPK* of cotton, *Arabidopsis*, and rice has not yet been determined. If *AtCIPK23* could interact with *AtCBL1* and *AtCBL9* simultaneously and respond to low potassium stress, and *AtCIPK23* and *GrCIPK30/GaCIPK16* are homologous gene pairs, then whether *GrCIPK30* or *GaCIPK16* would also simultaneously interact with CBL remains to be studied along with the mechanism of their interaction.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We would like to thank Dr. Cairui Lu for help in data analysis. Research supported by grants from the National High-tech R&D Program ("863" Program) (Grant #2011AA10A102).

REFERENCES

Albrecht V, Ritz O, Linder S, Harter K, et al. (2001). The NAF domain defines a novel protein-protein interaction module conserved in Ca^{2+}-regulated kinases. *EMBO J.* 20: 1051-1063. http://dx.doi.org/10.1093/emboj/20.5.1051

Assmann SM and Wang XQ (2001). From milliseconds to millions of years: guard cells and environmental responses. *Curr. Opin. Plant Biol.* 4: 421-428. http://dx.doi.org/10.1016/S1369-5266(00)00195-3

Carra A, Gambino G and Schubert A (2007). A cetyltrimethylammonium bromide-based method to extract low-molecular-weight RNA from polysaccharide-rich plant tissues. *Anal. Biochem.* 360: 318-320. http://dx.doi.org/10.1016/j.ab.2006.09.022

Chae MJ, Lee JS, Nam MH, Cho K, et al. (2007). A rice dehydration-inducible SNF1-related protein kinase 2 phosphorylates an abscisic acid responsive element-binding factor and associates with ABA signaling. *Plant Mol. Biol.* 63: 151-169. http://dx.doi.org/10.1007/s11103-006-9079-x

Chen L, Ren F, Zhou L, Wang QQ, et al. (2012). The Brassica napus calcineurin B-Like 1/CBL-interacting protein kinase 6 (*CBL1/CIPK6*) component is involved in the plant response to abiotic stress and ABA signalling. *J. Exp. Bot.* 63: 6211-6222. http://dx.doi.org/10.1093/jxb/ers273

Chen L, Wang QQ, Zhou L, Ren F, et al. (2013). Arabidopsis CBL-interacting protein kinase (*CIPK6*) is involved in plant response to salt/osmotic stress and ABA. *Mol. Biol. Rep.* 40: 4759-4767. http://dx.doi.org/10.1007/s11033-013-2572-9

Chen X, Gu Z, Xin D, Hao L, et al. (2011). Identification and characterization of putative *CIPK* genes in maize. *J. Genet. Genomics* 38: 77-87. http://dx.doi.org/10.1016/j.jcg.2011.01.005

Chothia C, Gough J, Vogel C and Teichmann SA (2003). Evolution of the protein repertoire. *Science* 300: 1701-1703. http://dx.doi.org/10.1126/science.1085371

Flagel LE and Wendel JF (2009). Gene duplication and evolutionary novelty in plants. *New Phytol.* 183: 557-564. http://dx.doi.org/10.1111/j.1469-8137.2009.02923.x

Halfter U, Ishitani M and Zhu JK (2000). The *Arabidopsis* SOS2 protein kinase physically interacts with and is activated by the calcium-binding protein SOS3. *Proc. Natl. Acad. Sci. USA* 97: 3735-3740. http://dx.doi.org/10.1073/pnas.97.7.3735

Harper JF (2001). Dissecting calcium oscillators in plant cells. *Trends Plant Sci.* 6: 395-397. http://dx.doi.org/10.1016/S1360-1385(01)02023-4

He DH, Lei ZP, Tang BS, Xing HY, et al. (2015). Identification and analysis of the *TIFY* gene family in *Gossypium raimondii*. *Genet. Mol. Res.* 14: 10119-10138. http://dx.doi.org/10.4238/2015.August.21.19

He L, Yang X, Wang L, Zhu L, et al. (2013). Molecular cloning and functional characterization of a novel cotton CBL-interacting protein kinase gene (*GhCIPK6*) reveals its involvement in multiple abiotic stress tolerance in transgenic plants. *Biochem. Biophys. Res. Commun.* 435: 209-215. http://dx.doi.org/10.1016/j.bbrc.2013.04.080

Huang C, Ding S, Zhang H, Du H, et al. (2011). *CIPK7* is involved in cold response by interacting with CBL1 in Arabidopsis thaliana. *Plant Sci.* 181: 57-64. http://dx.doi.org/10.1016/j.plantsci.2011.03.011

Huertas R, Olías R, Eljakaoui Z, Gálvez FJ, et al. (2012). Overexpression of SlSOS2 (*SlCIPK24*) confers salt tolerance to transgenic tomato. *Plant Cell Environ.* 35: 1467-1482. http://dx.doi.org/10.1111/j.1365-3040.2012.02504.x

Iqbal K, Azhar FM, Khan IA, Ehsan-Ullah (2011). Variability for Drought Tolerance in Cotton (*Gossypium hirsutum*) and its Genetic Basis. *Int. J. Agric. Biol.* 13: 61-66.

Kolukisaoglu U, Weinl S, Blazevic D, Batistic O, et al. (2004). Calcium sensors and their interacting protein kinases: genomics of the *Arabidopsis* and rice CBL-CIPK signaling networks. *Plant Physiol.* 134: 43-58. http://dx.doi.org/10.1104/pp.103.033068

Lecharny A, Boudet N, Gy I, Aubourg S, et al. (2003). Introns in, introns out in plant gene families: a genomic approach of the dynamics of gene structure. *J. Struct. Funct. Genomics* 3: 111-116. http://dx.doi.org/10.1023/A:1022614001371

Li F, Fan G, Wang K, Sun F, et al. (2014). Genome sequence of the cultivated cotton *Gossypium arboreum*. *Nat. Genet.* 46: 567-572. http://dx.doi.org/10.1038/ng.2987

Li LB, Zhang YR, Liu KC, Ni ZF, et al. (2010). Identification and Bioinformatics Analysis of *SnRK2* and *CIPK* Family Genes in Sorghum. *Agric. Sci. China* 9: 19-30. http://dx.doi.org/10.1016/S1671-2927(09)60063-8

Long M, Rosenberg C and Gilbert W (1995). Intron phase correlations and the evolution of the intron/exon structure of genes. *Proc. Natl. Acad. Sci. USA* 92: 12495-12499. http://dx.doi.org/10.1073/pnas.92.26.12495

Mahajan S, Sopory SK and Tuteja N (2006). Cloning and characterization of CBL-CIPK signaling components from a legume (*Pisa sativum*). *FEBS J.* 273: 907-925.

Mortazavi A, Williams BA, McCue K, Schaeffer L, et al. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5: 621-628. http://dx.doi.org/10.1038/nmeth.1226

Pandey GK, Cheong YH, Kim BG, Grant JJ, et al. (2007). *CIPK9*: a calcium sensor-interacting protein kinase required for low-potassium tolerance in Arabidopsis. *Cell Res.* 17: 411-421. http://dx.doi.org/10.1038/cr.2007.39

Paterson AH, Wendel JF, Gundlach H, Guo H, et al. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. *Nature* 492: 423-427. http://dx.doi.org/10.1038/nature11798

Roy SJ, Huang W, Wang XJ, Evrard A, et al. (2013). A novel protein kinase involved in Na(+) exclusion revealed from positional cloning. *Plant Cell Environ.* 36: 553-568. http://dx.doi.org/10.1111/j.1365-3040.2012.02595.x

Sanders D, Pelloux J, Brownlee C and Harper JF (2002). Calcium at the crossroads of signaling. *Plant Cell* 14 (Suppl): S401-S417.

Schauser L, Wieloch W and Stougaard J (2005). Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*. *J. Mol. Evol.* 60: 229-237. http://dx.doi.org/10.1007/s00239-004-0144-2

Schwachtje J, Minchin PEH, Jahnke S, van Dongen JT, et al. (2006). SNF1-related kinases allow plants to tolerate herbivory by allocating carbon to roots. *Proc. Natl. Acad. Sci. USA* 103: 12935-12940. http://dx.doi.org/10.1073/pnas.0602316103

Tang RJ, Liu H, Bao Y, Lv QD, et al. (2010). The woody plant poplar has a functionally conserved salt overly sensitive pathway in response to salinity stress. *Plant Mol. Biol.* 74: 367-380. http://dx.doi.org/10.1007/s11103-010-9680-x

Tripathi V, Parasuraman B, Laxmi A and Chattopadhyay D (2009). CIPK6, a CBL-interacting protein kinase is required for development and salt tolerance in plants. *Plant J.* 58: 778-790. http://dx.doi.org/10.1111/j.1365-313X.2009.03812.x

Wang K, Wang Z, Li F, Ye W, et al. (2012). The draft genome of a diploid cotton *Gossypium raimondii*. *Nat. Genet.* 44: 1098-1103. http://dx.doi.org/10.1038/ng.2371

Wang QQ, Liu F, Chen XS, Ma XJ, et al. (2010). Transcriptome profiling of early developing cotton fiber by deep-sequencing reveals significantly differential expression of genes in a fuzzless/lintless mutant. *Genomics* 96: 369-376. http://dx.doi.org/10.1016/j.ygeno.2010.08.009

Wei KF, Wang YM and Xie DX (2014). Identification and expression profile analysis of the protein kinase gene superfamily in maize development. *Mol. Breed.* 33: 155-172. http://dx.doi.org/10.1007/s11032-013-9941-x

Weinl S and Kudla J (2009). The CBL-CIPK $Ca^{(2+)}$-decoding signaling network: function and perspectives. *New Phytol.* 184: 517-528. http://dx.doi.org/10.1111/j.1469-8137.2009.02938.x

Xiang Y, Huang Y and Xiong L (2007). Characterization of stress-responsive *CIPK* genes in rice for stress tolerance improvement. *Plant Physiol.* 144: 1416-1428. http://dx.doi.org/10.1104/pp.107.101295

Xu J, Li HD, Chen LQ, Wang Y, et al. (2006). A protein kinase, interacting with two calcineurin B-like proteins, regulates K^+ transporter AKT1 in *Arabidopsis*. *Cell* 125: 1347-1360. http://dx.doi.org/10.1016/j.cell.2006.06.011

Yin Z, Wang J, Wang D, Fan W, et al. (2013). The *MAPKKK* gene family in *Gossypium raimondii*: genome-wide identification, classification and expression analysis. *Int. J. Mol. Sci.* 14: 18740-18757. http://dx.doi.org/10.3390/ijms140918740

Supplementary material

Table S1. The primers of fluorescence quantitative real-time PCR of the *CIPK* gene family of diploid cotton under different abiotic stress conditions.

Table S2. Gene pairs in D genome and A genome of the *CIPK* gene family.

Table S3. The *CIPK* homologous gene pairs between cotton and other specices.

Table S4. The specific expression of the *CIPK* gene family in leaves, stems and roots of *Gossypium raimondii* and *Gossypium arboreum* L. under different stresses.