Genome-wide identification and functional characterization of *CDPK* gene family reveal their involvement in response to drought stress in *Gossypium barbadense*

Guangzhen Shi and Xinxia Zhu

Key Laboratory of Oasis Town and Mountain-basin System Ecology of Xinjiang Production and Construction Corps, College of Life Sciences, Shihezi University, Shihezi, Xinjiang, China

ABSTRACT

Background. Calcium dependent protein kinases (CDPKs) are a class of important calcium signal sensing response proteins, which play an important regulatory role in response to abiotic stress. However, researchers have not been excavated CDPKs' role in drought in sea-island cotton(*Gossypium barbadense* L. 'H7124').

Results. Eighty-four *CDPK* genes have been identified in *G. barbadense*. These *GbCDPK* genes are unevenly distributed on 26 chromosomes, and segmental duplication is the significant way for the extension of CDPK family. Also, members within the same sub-family share a similar gene structure and motif composition. There are a large number of cis-elements involved in plant growth and response to stresses in the promoter regions of GbCDPKs. Additionally, these GbCDPKs show differential expression patterns in cotton tissues. The transcription levels of most genes were markedly altered in cotton under heat, cold, salt and PEG treatments, while the expressions of some GbCDPKs were induced in cotton under drought stress. Among these drought-induced genes, we selected *GbCDPK32*, *GbCDPK68*, *GbCDPK74*, *GbCDPK80* and *GbCDPK83* for further functional characterization by virus-induced gene silencing (VIGS) method.

Conclusions. In conclusion, the principal findings of this prospective study are that CDPKs were associated with drought. These findings provide a solid foundation for the development of future molecular mechanism in sea-island cotton.

Subjects Agricultural Science, Bioinformatics, Molecular Biology, Plant Science
Keywords Identification, CDPKs, Drought, Stress, *Gossypium barbadense*

INTRODUCTION

Cotton is a significant financial and oil crop in the world, which provides the most commonly natural fiber materials for cotton textile industry. Hence, cotton has been developed for a huge scope around the world. Xinjiang is the largest cotton-growing region in China and even in the world. Sea-island cotton is the second-largest cotton planted in the world, which has shown in comparable superiority for its highly desirable fiber properties (*Yu et al., 2021*), which is grown in large areas in southern Xinjiang, where it is arid, with long sunlight and little rainfall. The growing area of sea-island cotton is limited due to the

relatively low lint yield and narrow adaptation. Drought is one of the most serious climatic disasters affecting cotton quality and production (*Ault, 2020*). This is a problem that needs to be solved urgently.

Calcium dependent protein kinases are a class of Ser/Thr type protein kinases, which are usually composed of four core domains: N-terminal variable domain, protein kinase domain, autoinhibitory domain and C-terminal calmodulin like domain (*Cheng et al., 2002*). Plants sense environmental stimuli and transmit extracellular signals into intracellular through typical mechanisms and cause a series of responses. The C-terminal calmodulin like domain is composed of a pair of helix loop EF-hands structure. The combination of Ca^{2+} and EF-hands structure can cause the change of the spatial structure of CDPKs, which leads to the release of the inhibition of the self inhibitory domain, and promotes the recovery of the protein kinase activity of CDPKs, phosphorylates the downstream regulatory factors, and transmits the signal to the downstream regulatory network (*Shi et al., 2018*).

When plants are faced with abiotic stresses, CDPKs can respond to abiotic stresses and improve plant resistance to abiotic stresses by regulating the specific expression of genes. Overexpression of *AtCPK4*, *AtCPK8*, *AtCPK10* and *AtCPK11* can significantly improve the drought resistance of transgenic plants (*Zou et al., 2010*; *Zhu et al., 2007*; *Zou et al., 2015*). It is reported that *OsCDPK4* plays an important role in salt tolerance and drought stress of rice (*Campo et al., 2014*). Overexpression of *OsCDPK9* play an active role in drought stress tolerance (*Wei et al., 2014*). Overexpression of tobacco *ZoCDPK1* gene can affect salt and drought stress (*Vivek et al., 2017*). Overexpression of *ZmCPK4* in the transgenic *Arabidopsis* enhanced drought stress tolerance (*Jiang et al., 2013*).

Although CDPKs genome-wide identification and functional studies have been carried out in *Arabidopsis* (34) (*Hrabak et al., 2003*), rice (31) (*Ray et al., 2007*) and maize (31) (*Kong et al., 2013*), there is no report on *CDPK* gene family in *G. barbadense*. Based on *G. barbadense* genomes, we used bioinformatics to identify the members of *CDPK* gene family and predict the molecular structure, physiological characteristics of the protein. Analyzing the chromosome location, evolution and classification, the expression profiles of the *CDPK* gene. Moreover, we further investigated the function of the selected CDPKs in *G. barbadense* defense against drought stress by using virus-induced gene silencing (VIGS) technology. The results of this experiment provide an important theoretical basis for further understanding of the evolution and function of the *CDPK* and the cultivation of drought-tolerant cotton varieties in *G. barbadense*.

MATERIALS & METHODS

Identification of the CDPK family genes in *Gossypium barbadense*

The *G. barbadense* genome data comes from https://cottonfgd.org/ (*Zhu et al., 2017*), The *Arabidopsis* CDPK protein sequences of family members were downloaded from the *Arabidopsis* genome database TAIR website (https://www.arabidopsis.org/). Using Pfam online database (http://pfam.xfam.org/) to download the *CDPK* gene seed file EF-hand (Pfam ID: PF00036) and serine/threonine protein kinase (Pfam ID: PF07714). The

hmmsearch program of HMMER3.0 was used to identify protein sequences containing CDPKs conserved domains. All CDPK protein sequences were extracted NCBI CDD and Pfam online database were used to verify the conservative domain. Finally, all members of *CDPK* gene family were obtained and named according to their different positions on chromosomes in *G. barbadense*.

Molecular structure and physicochemical properties

The Protparam tool (https://web.expasy.org/protparam/) was used to predict the molecular weight, isoelectric point and other relevant information of the CDPK protein. Using the online tool myristoylation (https://web.expasy.org/myristoylator/) predict *N*-myristoylation, and palmitoylation prediction tools is CSS-plam. Subcellular localization of the CDPK proteins were predicted using an online tool wolf POSRT (https://www.genscript.com/psort.html).

Phylogenetic tree, gene structure, conserved motif, and promoter region analyses

The ClustalX software was used to perform multiple sequence alignments on the CDPK proteins of *G. barbadense* and *Arabidopsis*, and the phylogenetic tree was constructed by MEGAX software: the Neighbor-Joining algorithm was used, and the Bootstrap repeat value was 1,000 times (*Kumar et al., 2018*). The online GSDS tool (http://gsds.cbi.pku.edu.cn) (*Hu et al., 2015*) was used to visualize the exon-intron structures of the CDPK genes. Using MEME (http://meme-suite.org/tools/meme) (*Bailey et al., 2009*) online software to analyze the conserved motifs of *G. barbadense* CDPK, the parameters were set as the maximum number of motifs found was 10, and the longest motif length was 100 nt (Nucleotide). The 1,500 bp DNA sequence upstream was extracted, and the PlantCARE (*Lescot et al., 2002*) database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to predict the possible cis-acting elements.

Chromosome location and gene replication

The information of *CDPK* gene family chromosome physical location was extracted from *G. barbadense* genome annotation file, and the chromosome physical distribution map of the genes was drawn with MapInspect. The definition of gene duplication is as follows: (1) the sequence length is greater than 80% of the gene; (2) the identity of alignment region is greater than 80%; (3) only one repeat event is calculated for closely linked genes. According to the chromosomal location of *CDPK* gene, tandem duplication and segment duplication can be identified. The KaKs_Calculator2.0 of Tbtools (*Chen et al., 2020*) was used to calculate the non-synonymous mutation rate (Ka) and the synonymous mutation rate (Ks) of the replication gene pair in *G. barbadense*.

Expression profile analysis, RNA extraction and real-time PCR

The transcriptome data of *GbCDPK* genes in the tissues and the abiotic stresses were from Sequence read archive, SRA, and the accession number was PRJNA274882. The RPKM value (the reads per kilobase of transcript per million mapped reads) represents the abundance of gene expression at the transcription level. Using Tbtools (*Chen et al., 2020*)

map the gene expression profiles of different tissues and different abiotic stresses. Total RNA is extracted with the plant total RNA extraction kit (TIANGEN, Beijing, China). RNA reverse transcription adopts HiScript II QRT SuperMix for qPCR (R222-01), a product of Nanjing Novazin. The cDNA was used as a template to perform RT-PCR and qPCR experiments. The internal reference gene is the ubiquitinated protein Ubiquitin7 (*UBQ7*), and the RT-PCR reaction uses 1.0% agarose gel to detect the amplification results. The quantitative instrument uses Roche products, and the calculated Ct value is used to calculate the relative expression of the gene. The calculation method is $2^{-\Delta\Delta Ct}$ (*Pfaffl, 2001*). Each gene amplification result contains at least 3 technical replicates and 3 biological replicates. The amplification system and procedure are shown in Appendix A and Appendix B.

Experiment treatment

The *G. barbadense* was grown for 4 weeks in a high humidity and high nutrient environment (nutritional soil: vermiculite: perlite = 3:1:1) at 28 °C, and then treated with abiotic stress to well growing cotton seedlings. Cold stress and high temperature stress are moved plants to incubators at 4 °C and 37 °C, respectively. For salt stress, seedlings are irrigated with 200 mM NaCl. Seedlings are irrigated with 20% PEG6000 for drought stress. One hour after the start of the treatment, the second true leaf of the treated cotton seedling was harvested and stored in a refrigerator at −80 °C. Each treatment included 3 seedlings, and each treatment had three experimental replicates.

VIGS experiment in *G. barbadense*

The silenced fragments of each *CDPK* gene were amplified from cotton and inserted into VIGS vector (TRV-00) to generate *CDPK* silent constructs. TRV-00 without the silent segment was used as control. Consequently, the vector was transformed into the GV3101. GV3101 contained the same amount of TRV-00 and TRV-*CDPK*, and infiltrated into the cotyledons of 14-days-old cotton seedlings by syringe infiltration method. Two weeks after inoculation, the silenced cotton seedlings were used for gene expression analysis and identification PEG treatment.

Methods of physiological index

RWC was measured by using the formula: $\text{RWC}(\%) = \frac{\text{FW}-\text{DW}}{\text{TW}-\text{DW}} \times 100$. Where FW, stands for fresh weight TW for turgid weight and DW for dry weight (*Loutfy et al., 2012*). *G. barbadense* leaves were soaked in 30 mL deionized water at 25 °C and then shaken for 30 min. Heated the solution to 100 °C for 30 min and then at room temperature. Measure REL with conductivity meter (*Yang et al., 2012*). The content of MDA, Pro, soluble sugar measured using the fresh sample. The amount of MDA was measured by the thiobarbituric acid method (*Heath & Packer, 1968*). The content of Pro was determined by the Sulfosalicylic acid indandione method (*Bates, Waldren & Teare, 1973*). Soluble sugar content was measured as described by the anthrone method and determined at 620 nm by a spectrophotometer (*Fales, 1951*). Water loss rate $(\%) = \frac{W1-W2}{W1} \times 100$. $W1$ stands for weight of detached blade before water loss. $W2$ weight of detached blade after 1, 2, 3, 4 and 5 h of water loss (*Czyczyło-Mysza et al., 2018*). All physiological results were repeated three

times and a one-way analysis of variance (ANOVA) was performed with Prism to detect significant differences between both treatments ($P < 0.05$, $P < 0.01$).

RESULTS

Identification and proteins characteristic of CDPKs in *Gossypium barbadense*

CDPK protein sequences of *Arabidopsis* were provided as direct queries against the *G. barbadense* genome to perform BLASTP and hidden Markov Model (HMM) to identify the *GbCDPK* gene family members. After NCBI-CDD and Pfam software database verification, 84 *CDPK* genes were identified in *G. barbadense* genome (Table S1).

The biochemical characteristics of the *CDPK* gene, such as amino acid sequence, molecular weight and isoelectric point were analyzed by the ProtParamtool online software. The results show that the *CDPK* genes are conservative. The longest CDPK protein encodes 648 amino acids, and the shortest only encodes 155 amino acids. The molecular weight of GbCDPKs varies 17.99 KDa to 71.854 KDa, and the theoretical isoelectric point ranges from 4.313 to 9.481. Multiple sequence alignment analysis indicated that all of GbCDPKs have four core domains. In addition, all the GbCDPKs have EF-hands motifs, which allows binding of Ca^{2+}. It reported that 71 GbCDPK proteins contain four EF-hands, nine proteins (GbCDPK25, GbCDPK26, GbCDPK39, GbCDPK49, GbCDPK53, GbCDPK67, GbCDPK68, GbCDPK77, GbCDPK81) have three EF-hands, and remaining four proteins (GbCDPK22, GbCDPK38, GbCDPK61, GbCDPK64) only contains 2 EF-hands. Our bioinformatic analysis revealed that 40 members harbored potential *N*-myristoylation sites, and 74 GbCDPK proteins contained S-palmitoylation sites (Table S1). Using TMHMMServerv2.0 prediction, it was found that none of the 84 CDPK proteins had a transmembrane structure. Using Wolf POSRT tool to predict subcellular localization, it was found that 22 CDPK proteins are in the chloroplast, 33 CDPK proteins are in the cytoplasm, eight are in the nucleus, and the others are in the mitochondria, extracellular, vacuole, endoplasmic reticulum, and peroxisome.

Phylogenetic analysis of GbCDPKs

To assess the phylogenetic relationship of CDPK protein in *G. barbadense*. A total of 118 CDPK proteins from *G. barbadense* and *Arabidopsis* were used to construct the phylogenetic tree by adjacency method (Fig. 1). According to the results of phylogenetic tree clustering, 84 *CDPK* genes can be divided into four subfamilies. Referring to the research results of *Arabidopsis*, four subfamilies were named as class I, class II, class III, and class IV. Further analysis of the clustering results of CDPK members revealed that the gene structures in the homologous gene pair were similar. The number of subgroup II proteins was the largest, including 30 *G. barbadense* and 10 *Arabidopsis* CDPK proteins, accounting for 35.71% and 29.41% of the total, respectively. The first subgroup includes 27 CDPK proteins from *G. barbadense* and 10 CDPK proteins from *Arabidopsis*. The third subgroup includes 19 CDPK proteins from *G. barbadense* and eight CDPK proteins from *Arabidopsis*. The fourth subgroup has the least number, including only eight CDPK proteins from *G. barbadense* and 3 CDPK proteins from *Arabidopsis*, indicating that it's far from each other.
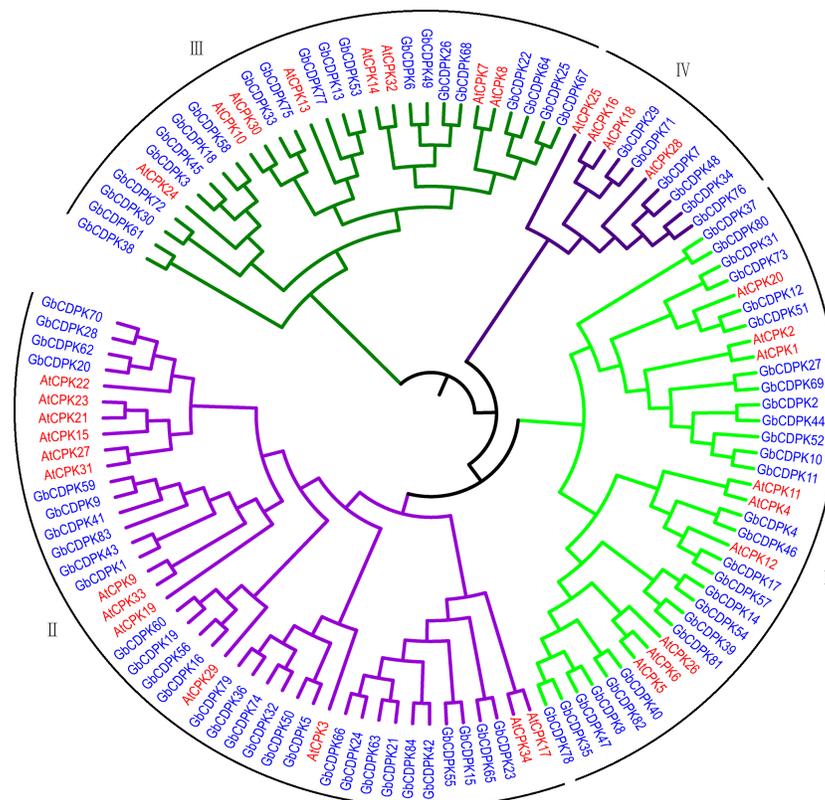
Figure 1 Evolutionary relationship of CDPK Family in *G. barbadense.*

Full-size ⬜ DOI: 10.7717/peerj.12883/fig-1

Chromosomal distribution and gene duplication

To investigate the chromosomal locations, chromosomal maps were drawn for the 84 *CDPK* genes from *G. barbadense* by MapInspect software. All 84 *GbCDPK* genes are mapped on 13 pairs of chromosomes (Fig. 2). Furthermore, cotton genome includes A-subgroup and D-subgroup that all contain 42 GbCDPKs, respectively, and the duplication events may illuminate the mechanism about the expansion of *GbCDPK* gene family. There are six genes on chromosome 4 and 13 of A-subgroup, five genes on chromosome 2 of A-subgroup and 5 and 13 of D-subgroup, respectively. In contrast, there is only one gene on chromosome 3 and 8 of A-subgroup and chromosome 8 of D-subgroup, and there are 2–4 genes on other chromosomes. These results indicated that *CDPK* genes were widely distributed in *G. barbadense*. In order to speculate the possible relationship between *GbCDPK* genes and potential gene replication in the *G. barbadense* genomes, we analyzed the occurrence of tandem duplication and large-scale segmental replication during the evolution of the gene family (Table S2). Among 84 genes, 41 pairs of homologous genes were found in phylogenetic tree (Fig. 1). There were 40 pairs of segmental replications and one pair of tandem replications (GbCDPK10/GbCDPK11).

To study the evolutionary replication relationship of genes in *G. barbadense* on chromosome segments, MCscanx software was used to detect the replication genes.

Figure 2 Distribution of CDPK gene family on chromosomes.
Full-size ▣ DOI: 10.7717/peerj.12883/fig-2

12, 16 and 60 pairs of whole genome replicates were detected among A, D and AD subgroups in *G. barbadense*, respectively (Fig. 2). KaKs_Calculator was used to perform Ka/Ks analysis replication gene pairs. It was found that the Ka/Ks of all gene pairs were less than 1, indicating that the *CDPK* gene of *G. barbadense* might have undergone strict purification and selection during the evolution process, implying that the replication gene is evolutionarily conserved, structurally stable, and might have consistent function.

Analysis of gene structure and conserved motifs

The *CDPK* genes were visualized for the gene structure and conserved motifs. The results showed that they have similar gene structure distribution patterns in the same subfamily. Most CDPKs of the first and the third subfamily have seven or eight exons, the CDPKs of the second subfamily have eight or nine exons, and the fourth subfamily has 12 exons. Most CDPKs of the family have 12 exons (Fig. 3). Generally, the structural conservation and similarity of the same type of genes are closely related to their evolutionary relationship, but there are some exceptions. *GbCDPK38* has only five exons and *GbCDPK61* has only three exons. The longest exon of most GbCDPKs is the first exon. However, there are still a small number of genes in the family members that have evolved significantly different structures. *GbCDPK39* and *GbCDPK81* are homologous genes from different sub-genomes, *GbCDPK39* has a typical gene structure of this group, while the first exon of *GbCDPK81* is relatively short. *GbCDPK10* and *GbCDPK11* are homologous genes of the same sub-genomes, but their gene structure is also different. *GbCDPK10* has typical gene structure characteristics, but the first exon of *GbCDPK11* is also shorter than that of
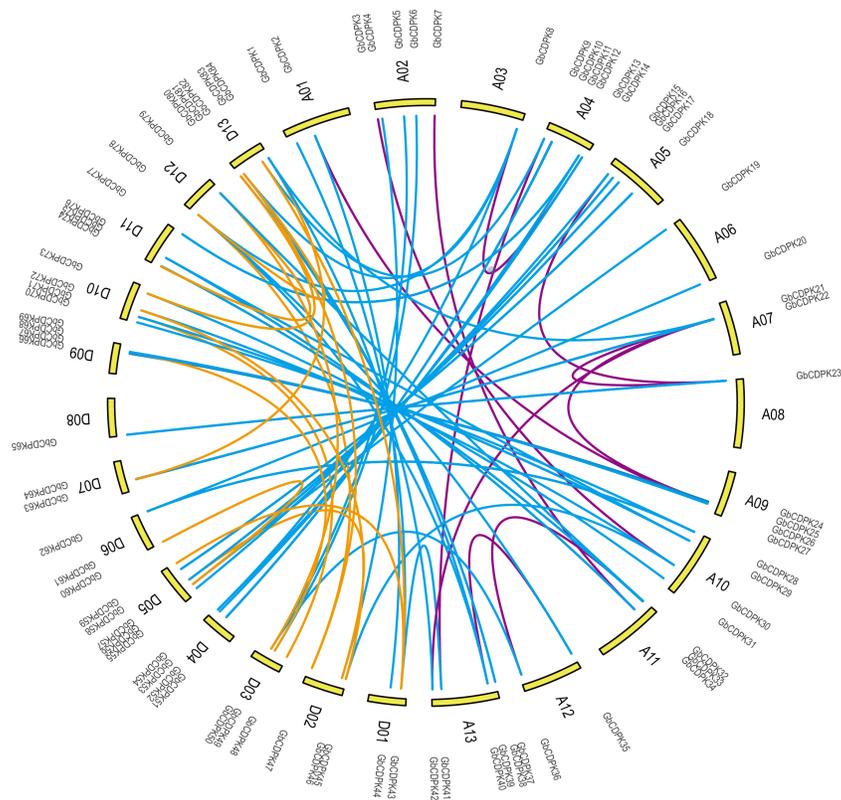
Figure 3 Whole genome duplication distribution of CDPK genes in *G. barbadense*.
Full-size ⬛ DOI: 10.7717/peerj.12883/fig-3

GbCDPK10. This structural difference is likely to have formed randomly during evolution. The changed gene structure may lead to changes in gene splicing sites, thereby generating new transcripts or non-functional gene products, and ultimately increasing the number of members of the cotton GbCDPKs gene family.

The MEME online software was used to conduct conservative analysis on the amino acid sequence of CDPK, and 10 conservative motifs were found (Fig. 4). These 10 conserved motifs exist in subgroups I, II, III and IV, but there are four genes (*GbCDPK11*, *GbCDPK14*, *GbCDPK38*, *GbCDPK61*) whose conservative motifs are distributed differently in the first subgroup. *GbCDPK11*, *GbCDPK14* contains the same conserved motifs, without Motif 2, Motif 3, Motif 5. While *GbCDPK38* has only three conserved motifs (Motif 8, Motif 9, Motif 10), *GbCDPK61* has two conserved motifs (Motif 8, Motif 10). However, in the same subfamily, *CDPK* has a similar number, type, and spatial distribution of motifs, demonstrating that the functions of *CDPK* genes in the same subfamily are similar.

Analysis of promoter cis-elements

To explore the regulatory mechanisms of *GbCDPK* genes, the 1.5kb sequence upstream of the start condon of the GbCDPKs was selected to analyze the constitutions of cis-regulatory elements. It was found that all *CDPK* genes contained CAAT-box conserved elements and TATA-box. In addition to these conserved elements, there are four types
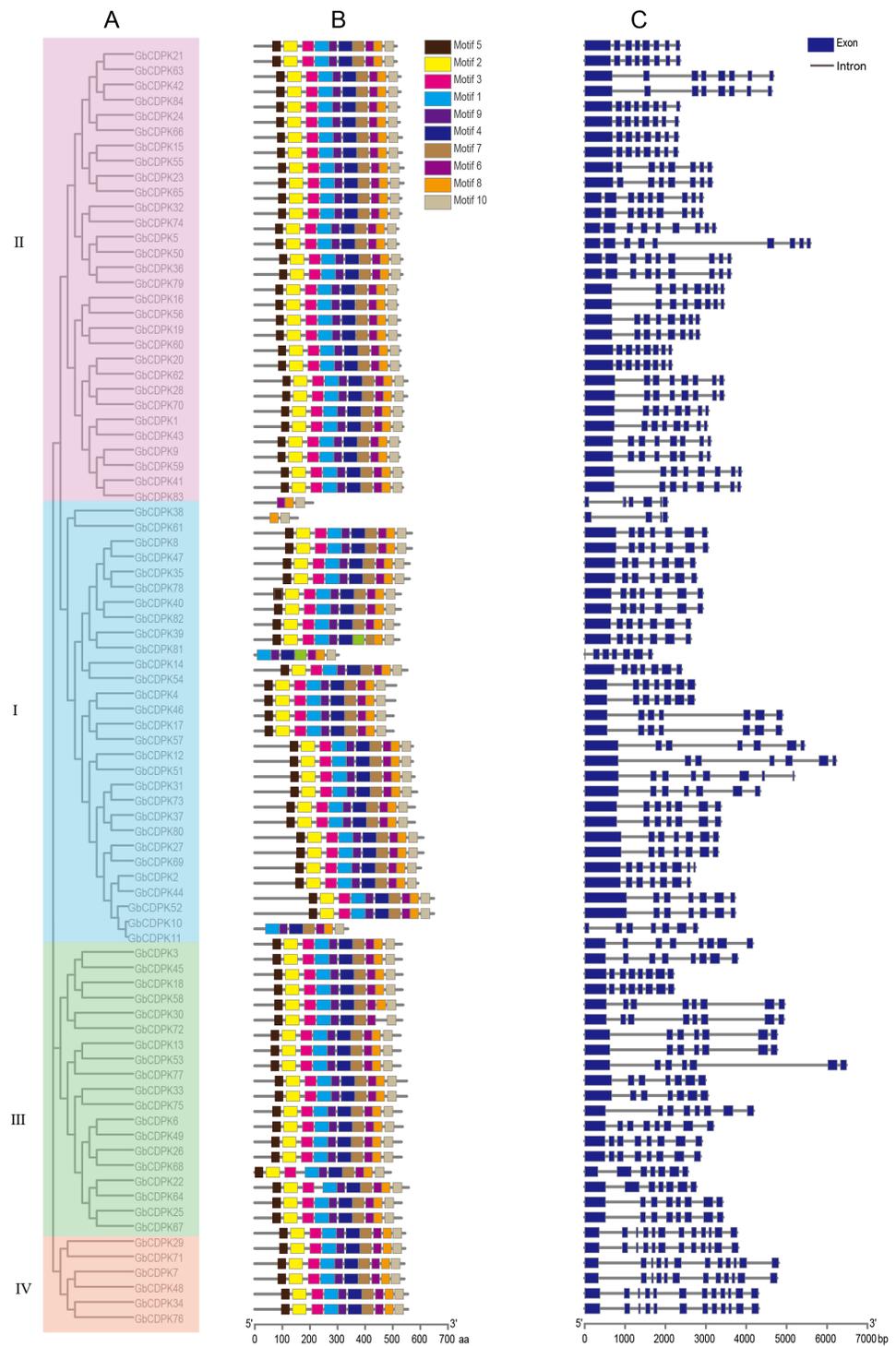
Figure 4 (A–C) Phylogenetic tree, conservative motif and gene structure in *G. barbadense*.
Full-size ⊡ DOI: 10.7717/peerj.12883/fig-4

of cis-regulatory elements in the promoter of *CDPK* gene: (1) Light regulatory elements, including Box 4, G-Box, TCCC-motif and AE-box, etc. (2) Plant growth and development regulatory elements, including CAT-box, GCN4_motif, O2-site, MBSI and circadian, etc. (3) Phytohormone response elements, including methyljasmonate response elements TGACG-motif and CGTCA-motif, (ABRE), gibberellin responsive element P-box, etc. (4) Stress response elements include LTR, MBS, ARE and mechanical damage response element W-box, etc.

Expression profile of different tissues and abiotic stress responses

To explore the possible biological functions of GbCDPKs, the transcriptome data of *G. barbadense* were used to analyze the expression characteristics of gene family in ovule and fiber at 0, 1, 3, 5, 10 and 20 days after anthesis, as well as in root, stem, leaf, receptacle, petal, stamen, calycle and sepals (Fig. 5A). The results showed that the *CDPK* gene has tissue-specific expression and spatio-temporal expression characteristics. According to the expression characteristics, the CDPK family can be divided into three types: the first group of genes are generally expressed less, but some genes were expressed in some or specific tissue. The second group was highly expressed in most tissues and developmental stages. The third group of genes are mainly expressed in vegetative organs or floral organs, but less expressed in ovule and fiber. Generally, homologous genes have similar expression patterns. Some homologous genes, such as *GbCDPK13*, *GbCDPK56* and *GbCDPK30*, *GbCDPK72* are highly expressed in the stamens. *GbCDPK28*, *GbCDPK70*; *GbCDPK33*, *GbCDPK75* and *GbCDPK38*, *GbCDPK80* are predominantly expressed in ovules. *GbCDPK1*, *GbCDPK43* and *GbCDPK25*, *GbCDPK67* are predominantly expressed in receptacles and stamens. The above results indicate that the *CDPK* gene family is widely involved in the growth and development of *G. barbadense*.

CDPKs can regulate the response of plants to abiotic stress. The transcriptome data was used to analyze the expression characteristics of the *CDPK* gene of *G. barbadense* under low temperature, high temperature, drought (PEG6000), and salt treatment (Fig. 5B). According to the expression characteristics shown in response to abiotic stress, the CDPK gene family can be divided into four groups: the first group most genes are low expressed under adversity stress. The second group all genes is low expressed. The three groups are all highly expressed. The fourth group most genes are highly expressed. In addition, homologous genes have similar expression profiles. For example, some homologous genes *GbCDPK31* and *GbCDPK73* are highly expressed under all stresses.

Expression under drought stress

The transcript levels of *CDPK* genes in cluster 1 and 2 were decreased under drought. In order to further understand the expression of *CDPK* genes at transcriptional level in *G. barbadense*. qRT-PCR was performed using RNA isolated from drought stress (Fig. 6). After drought, the expression of *GbCDPK4* showed a downward-increasing-decreasing trend, and the expression was the lowest at 6 h. *GbCDPK19* and *GbCDPK78* showed an overall upward-decreasing-increasing-decreasing trend. The former had the highest expression at 12 h, and the latter at 12 h. The expression level of *GbCDPK23* was the highest at 3
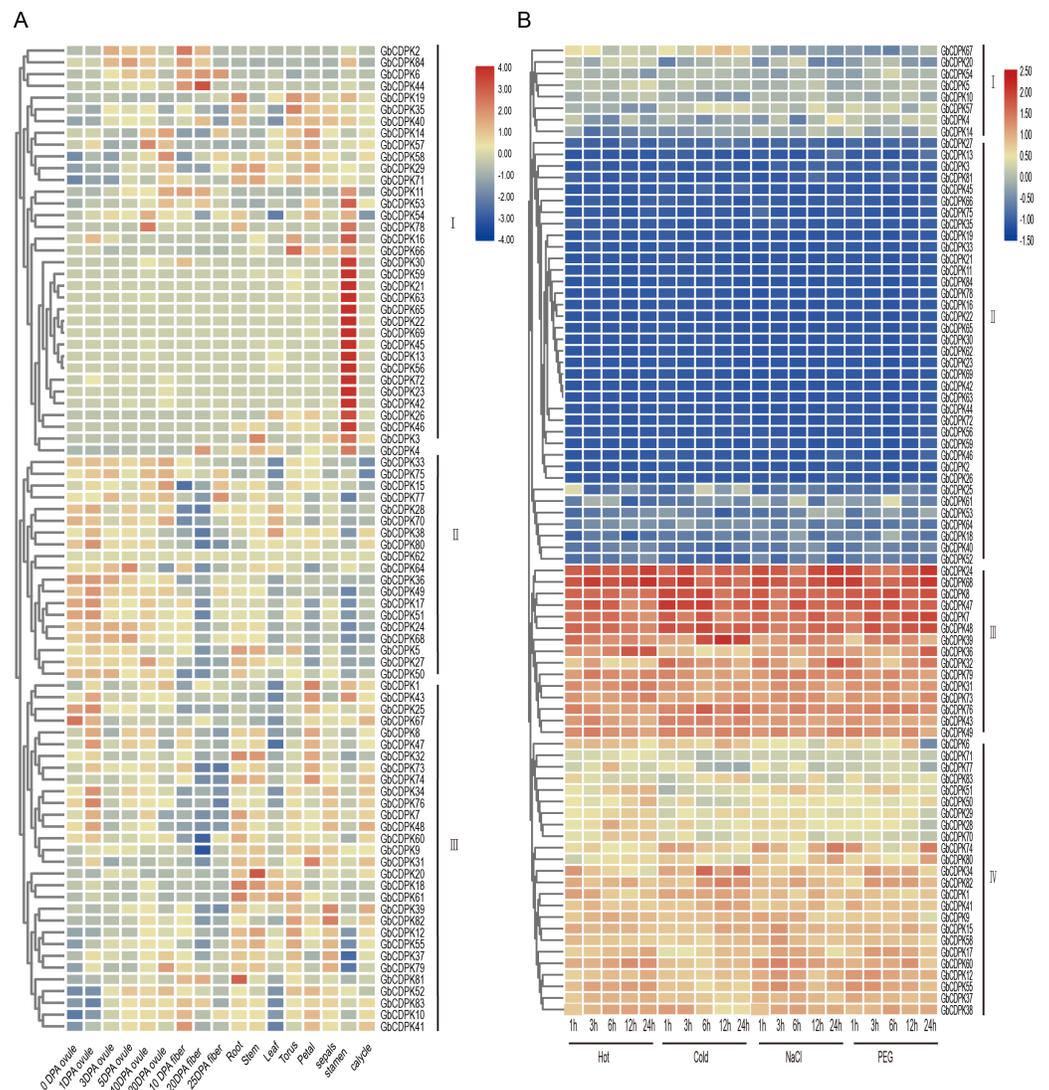
Figure 5 (A–B) Expression profile in different tissues and abiotic stress in GbCDPK.
Full-size ⊡ DOI: 10.7717/peerj.12883/fig-5

h. The expression level of *GbCDPK23* showed a trend of decrease-increasing-decreasing, and the expression level of 12 h was the highest. The expression levels of *GbCDPK44* and *GbCDPK66* showed a trend of increase-decrease-increasing, and the expression level of the former was the lowest at 6 h. The expression was highest at 24 h.

Silencing GbCDPKs compromises cotton resistance to drought stress

In order to further study the function of CDPKs drought tolerance in *G. barbadense*, we chose to design specific primers in the 3′-UTR region of CDPKs to construct VIGS vectors. The transcript levels of *CDPK* genes in cluster 3 and 4 were significantly increased under drought, with the highest peak observed all stage of the drought treatment. The empty vector (TRV-00) and the vector with target fragments (TRV-CDPKs) were activated and
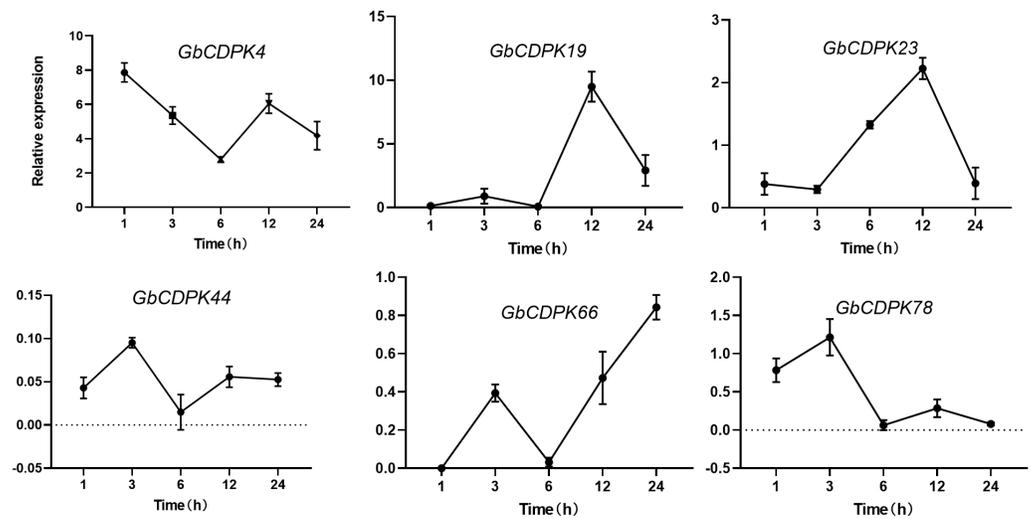
Figure 6 Abiotic stress expression of CDPK genes.

Full-size ⊠ DOI: 10.7717/peerj.12883/fig-6

propagated. Then the cotton cotyledons were injected into the 10 days seedlings for gene silencing. After two weeks of injection, when the positive seedlings were albino (Fig. 7A), indicating success of the VIGS experiment. The observed gene expression patterns may reflect their functions. The interference efficiency of the target gene was detected by qPCR technology. The results show that the *GbCDPKs* genes (*CDPK32*, *CDPK68*, *CDPK74*, *CDPK80* and *CDPK83*) in cotton plants injected with bacterial liquid containing the target gene fragments can indeed interfere effectively and specifically (Fig. 7B).

To determine the tolerance of GbCDPKs silenced plants under drought stress, leaf discs were made from the same parts of TRV-00 and TRV-GbCDPKs (TRV-*CDPK32*, TRV-*CDPK68*, TRV-*CDPK74*, TRV-*CDPK80* and TRV-*CDPK83*). The leaves were incubated in 0 and 20% PEG6000 solution for 4 days. The results showed that the leaf chlorosis and wilting of gene silenced TRV-*CDPK32*, TRV-*CDPK68*, TRV-*CDPK74*, TRV-*CDPK80* and TRV-*CDPK83* plants were significantly more chlorosis and wilting than that of the control (Fig. 7C).

In addition, we measured several physiological indicators including RWC, REL, water loss rate and MDA, proline and soluble sugar content under drought treatment and normal growth conditions. As shown in Figs. 7D–7H, in normal conditions, relative water content, electrical conductivity, proline and soluble sugar content of TRV-00 were higher than TRV-GbCDPKs. However, the TRV-00 of MDA was lower than silencing. Under the drought stress, the relative water content, proline and soluble sugar content were decreased in TRV-CDPKs plants, compared with the TRV-00. The relative electrical conductivity, MDA of TRV-00 and TRV-CDPKs both increased. From Fig. 7I, TRV-CDPKs was significantly higher than TRV-00, indicating that the drought resistance of plants decreased remarkably after silenced. In conclusion, the drought tolerance of the plants is weaker after silenced.
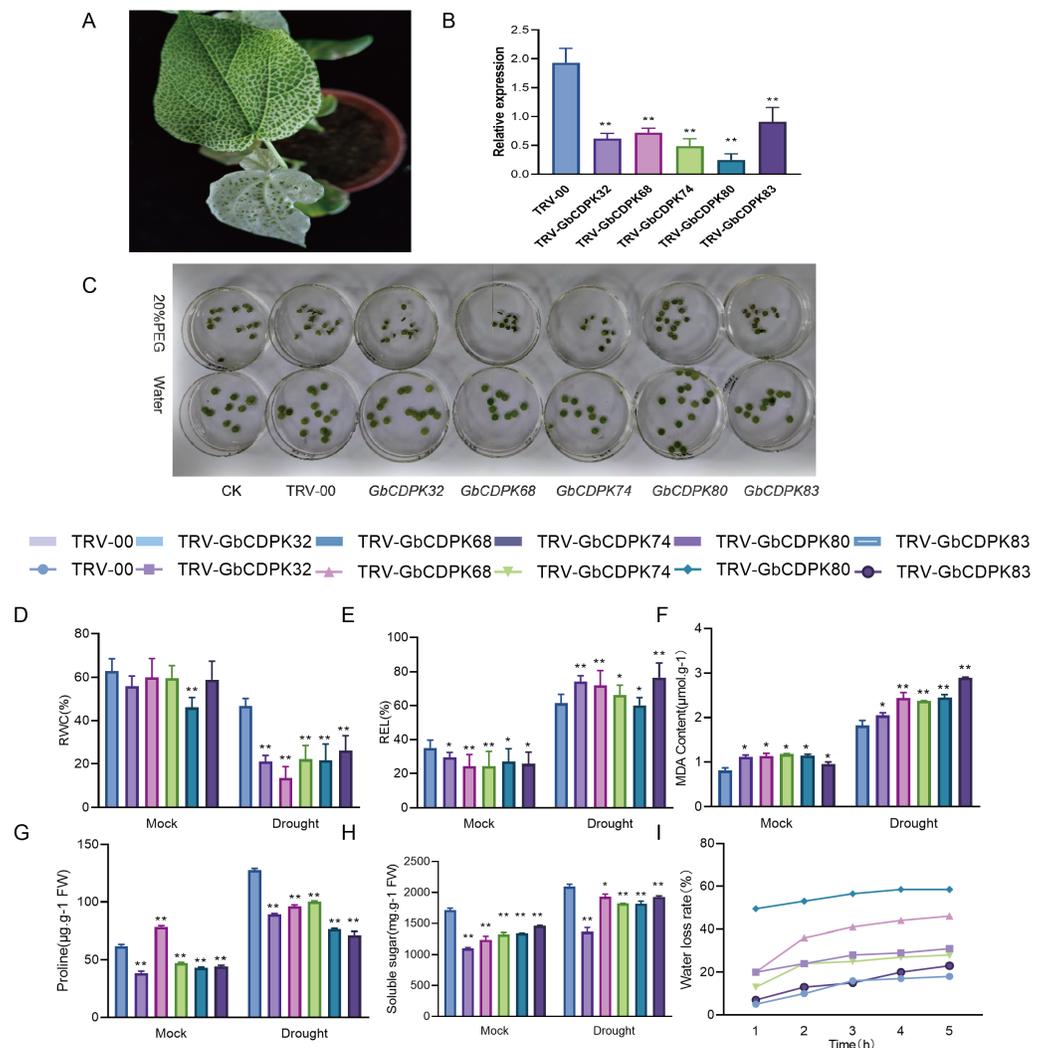
Figure 7 (A–I) Functional analysis of GbCDPKs in response to drought stress.

Full-size ⬛ DOI: 10.7717/peerj.12883/fig-7

DISCUSSION

In this study, we have identified 84 presumptive CDPK genes in *G. barbadense* (AADD). The members of CDPKs family are significantly more than those in other species. There are 41 pairs of homologous gene pairs in the *CDPK* gene family, and 40 pairs are segmental duplication between the A/D chromosome pairs. Our analysis showed that segmental duplication is a predominant driving force that contributed to the expansion of GbCDPKs. The parallel evolution characteristics of homologous gene pairs in A-subgroup and D-subgroup are significant. They may come from the same ancestor, but the remaining one pair exists in the form of tandem replication (*Liu et al., 2014*). Generally, the amplification of CDPKs family members in angiosperms mainly depends on large-scale gene rearrangement, that is, whole genome or large fragment gene replication events (*Hamel, Sheen & Séguin, 2014*). This may explain the fact that there are more members of

CDPKs family in *G. barbadense* than in *Arabidopsis*, rice and maize. Phylogenetic analysis indicated the GbCDPKs divided into four groups like *Arabidopsis*, and most CDPKs in the same group have similar gene structure, indicating the evolutionary and functional conservation of CDPKs. Generally, the orthologous genes have similar biological functions, while the paralogous genes have different biological functions. Therefore, the function of *GbCDPK* genes can be inferred from orthologous genes to provide scientific theoretical basis for subsequent functional studies (*Thornton & DeSalle, 2000*).

The specificity of signal transduction can be determined by the subcellular location of CDPK protein. It is predicted that GbCDPKs are in several cell compartments. It is also predicted that most GbCDPKs contain *N*-myristoylation and palmitoylation. The *CDPK* genes of *N*- myristoylation motif tends to be in the plasma membrane (*Martín & Busconi, 2000*; *Lu & Hrabak, 2013*) The myristoylation may be a part of the main signal transduction process that guides proteins to membrane binding sites. In addition, it is reported that another lipid-modified palmitoylation is necessary for the stability of membrane association (*Asai et al., 2013*). In this study, 40 genes have *N*-myristoylation and 75 genes have palmitoylation motifs. However, the association of CDPKs with membranes is complex and may be affected by other motifs. For example, in wheat (*Triticum aestivum*), *TaCPK3* and *TaCPK15* without myristoylation regions are associated with the membrane (*Li et al., 2008*), while in Maize (*Zea mays*), *N*-myristoylation patterns were found in *ZmCPK1* of the cytoplasm and nucleus (*Wang & Shao, 2013*).

Other studies have shown that the subcellular location of CDPKs can be limited in a single compartment or widely distributed in the whole cell. CDPKs have been found in plasma membrane, cytoplasm, nucleus, endoplasmic reticulum, mitochondria, chloroplast, oleosome, peroxisome and Golgi apparatus (*Asano et al., 2012a*). To explore the function of proteins in plant, subcellular localization is one of the essential research. Ca^{2+} pumps are known to be widely distributed on membranes (*Tanaka et al., 2004*), CDPK can rapidly respond to transient changes of Ca^{2+} signal in plants and specifically recognize substrates after phosphorylation. Various physiological responses were further triggered in signaling transduction cascades, which finally regulate plant growth and development, and in response to multiple stresses (*Shi et al., 2018*). We found that StCDPK7 protein with auto-phosphorylation activity could locate in the nucleus, membranes and interact with PAL protein, suggesting that the auto-phosphorylation activity of StCDPK7 protein played important roles in protein location and protein-protein interactions (*Fantino et al., 2017*). SiCDPK24 protein was proven to have a role in auto-phosphorylation, and may be important for its function. SiCDPK24-GFP localized to the plasma membrane and the nucleus (*Yu et al., 2018*). Therefore, the subcellular localization of *CDPK* genes are closely related to drought stress response.

CDPKs activated by transient changes in Ca^{2+} concentration in plants is essential for various biological processes (such as plant growth, development, regeneration, defense against biotic and abiotic stresses, etc.) (*Asano et al., 2012a*; *Ranty et al., 2016*). It has been reported that CDPKs are widely distributed in plant tissues, even in fruit, pollen tubes, germs, and guard cells (*Chang et al., 1995*; *Zhou, Fu & Yang, 2009*). The CDPKs protein family is huge, and its functions are also diverse (*Asano et al., 2012b*). We use the

transcriptome data to evaluate the expression pattern of cotton CDPKs in 17 different tissues. The results of tissue expression distribution indicated that the gene family of cotton CDPKs was widely distributed, implying that its function is diverse. CDPKs expressed in many tissues may participate in a variety of biological processes, some members specifically expressed in stamens may participate in reproductive development, and some may participate in stress response process. The specific genes involved in which biological process need to be further studied.

With the change of global ecological environment, drought has become one of the main environmental factors limiting crop quality and yield. In other plants, CDPKs was associated with drought tolerance, such as *StCDPK3* and *StCDPK23* and *PtrCDPK10* are all involved in drought stress (*Bi et al., 2021*; *Meng et al., 2020*). However, few people have studied the function of CDPKs in *G. barbadense.*

To further study the function of drought tolerance in GbCDPKs, VIGS technology was used to silence candidate *CDPK* genes to construct transient silence materials (*Senthil-Kumar & Mysore, 2014*). Our analysis indicated that CDPKs silencing plants were more sensitive to drought stress than control plants. The results of the study showed that the drought resistance ability of CDPKs silencing was significantly reduced. The content of proline and the MDA was higher than the control. On the contrary, the relative water content of the silenced plants was lower than that in the controls under drought stress. To maintain the stability of protoplast colloid, the proline is accumulated to increase the abiotic pressure of plant cells under drought stress (*Sallam et al., 2019*). In this study, drought stress induced accumulation of proline in cotton, and thereby enhanced abiotic stress adjustment ability. The drought sensitive phenotype of silencing plants to drought stress indicates that CDPKs play a regulatory role in improving cotton drought tolerance, but whether they participate in the same drought tolerance regulatory mechanism is still unknown, and further research is needed to prove it.

CONCLUSIONS

Eight-four *CDPK* genes were identified in *G. barbadense*. They all have the EF-hands structure and Ser/Thr protein kinase domain, none of which include the transmembrane structure. They were found to express in different organelles by subcellular localization. According to the phylogenetic tree, all CDPKs were divided into four groups, which were distributed on 13 pairs of chromosomes. The CDPKs are distributed in 13 tissues. In response to abiotic stress, it showed different expression patterns. The silence of some CDPKs severely destroyed the basic resistance of cotton to drought stress, indicating that they were involved in the resistance mechanism. This study deepens our understanding of the function of CDPK and provide important insights for further research on the molecular mechanism of *G. barbadense* to drought stress and to provide candidate target genes that improve drought adaptation.

 Abbreviations

CDPKs	Calcium dependent protein kinases

VIGS	virus-induced gene silencing
qRT-PCR	quantitative reverse transcription PCR
G. barbadense	*Gossypium barbadense*
RWC	Relative water content
REL	Relative electrical conductivity leakage
MDA	Malondialdehyde

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Guangzhen Shi performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Xinxia Zhu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:
The raw measurements are available in the Supplementary File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12883#supplemental-information.

REFERENCES

Asai S, Ichikawa T, Nomura H, Kobayashi M, Kamiyoshihara Y, Mori H, Kadota Y, Zipfel C, Jones JDG, Yoshioka H. 2013. The variable domain of a plant calcium-dependent protein kinase (CDPK) confers subcellular localization and substrate recognition for NADPH oxidase. *Journal of Biological Chemistry* **288(20)**:14332–14340 DOI 10.1074/jbc.M112.448910.

Asano T, Hayashi N, Kikuchi S, Ohsugi R. 2012a. CDPK-mediated abiotic stress signaling. *Plant Signaling & Behavior* **7**(7):817–821 DOI 10.4161/psb.20351.

Asano T, Hayashi N, Kobayashi M, Aoki N, Miyao A, Mitsuhara I, Ichikawa H, Komatsu S, Hirochika H, Kikuchi S, Ohsugi R. 2012b. A rice calcium-dependent protein kinase *OsCPK12* oppositely modulates salt-stress tolerance and blast disease resistance. *The Plant Journal* **69**(1):26–36 DOI 10.1111/j.1365-313X.2011.04766.x.

Ault TR. 2020. On the essentials of drought in a changing climate. *Science* **368**(6488):256–260 DOI 10.1126/science.aaz5492.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**(**Web Server issue**):W202–W208 DOI 10.1093/nar/gkp335.

Bates LS, Waldren RP, Teare ID. 1973. Rapid determination of free proline for water-stress studies. *Plant and Soil* **39**(1):205–207 DOI 10.1007/BF00018060.

Bi Z, Wang Y, Li P, Sun C, Qin T, Bai J. 2021. Evolution and expression analysis of *CDPK* genes under drought stress in two varieties of potato. *Biotechnology Letters* **43**(2):511–521 DOI 10.1007/s10529-020-03037-2.

Campo S, Baldrich P, Messeguer J, Lalanne E, Coca M, San SB. 2014. Overexpression of a calcium-dependent protein kinase confers salt and drought tolerance in rice by preventing membrane lipid peroxidation. *Plant Physiology* **165**(2):688–704 DOI 10.1104/pp.113.230268.

Chang A, Condron R, Neumann GM, Polya GM. 1995. Purification and characterization of a heat-stable wheat substrate for wheat embryo calcium-dependent protein kinase. *Biochimica et Biophysica Acta* **1244**(2–3):317–324 DOI 10.1016/0304-4165(95)00039-e.

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* **13**(8):1194–1202 DOI 10.1016/j.molp.2020.06.009.

Cheng SH, Willmann MR, Chen HC, Sheen J. 2002. Calcium signaling through protein kinases. The *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiology* **129**(2):469–485 DOI 10.1104/pp.005645.

Czyczyło-Mysza IM, Marcińska I, Skrzypek E, Bocianowski J, Dziurka K, Rančić D, Radošević R, Pekić-Quarrie S, Dodig D, Quarrie SA. 2018. Genetic analysis of water loss of excised leaves associated with drought tolerance in wheat. *PeerJ* **6**:e5063 DOI 10.7717/peerj.5063.

Fales FW. 1951. The assimilation and degradation of carbohydrates by yeast cells. *Journal of Biological Chemistry* **193**:113–124 DOI 10.1007/s10482-013-0100-5.

Fantino E, Segretin ME, Santin F, Mirkin FG, Ulloa RM. 2017. Analysis of the potato calcium-dependent protein kinase family and characterization of stcdpk7, a member induced upon infection with Phytophthora infestans. *Plant Cell Reports* **36**:1137–1157 DOI 10.1007/s00299-017-2144-x.

Hamel LP, Sheen J, Séguin A. 2014. Ancient signals: comparative genomics of green plant CDPKs. *Trends in Plant Science* **19**(2):79–89 DOI 10.1016/j.tplants.2013.10.009.

Heath RL, Packer L. 1968. Photoperoxidation in isolated chloroplasts. I—kinetics and stoichiometry of fatty acid peroxidation. *Archives of Biochemistry and Biophysics* **125(1)**:189–198 DOI 10.1016/0003-9861(68)90654-1.

Hrabak EM, Chan CW, Gribskov M, Harper JF, Choi JH, Halford N, Kudla J, Luan S, Nimmo HG, Sussman MR, Thomas M, Walker-Simmons K, Zhu JK, Harmon AC. 2003. The *Arabidopsis* CDPK-SnRK superfamily of protein kinases. *Plant Physiology* **132(2)**:666–680 DOI 10.1104/pp.102.011999.

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. 2015. GSDS 2.0: an upgraded gene features visualization server. *Bioinformatics* **31(8)**:1296–1297 DOI 10.1093/bioinformatics/btu817.

Jiang S, Zhang D, Wang L, Pan J, Liu Y, Kong X, Zhou Y, Li D. 2013. A maize calcium-dependent protein kinase gene, ZmCPK4, positively regulated abscisic acid signaling and enhanced drought stress tolerance in transgenic *Arabidopsis. Plant Physiology and Biochemistry* **71**:112–120 DOI 10.1016/j.plaphy.

Kong X, Lv W, Jiang S, Zhang D, Cai G, Pan J, Li D. 2013. Genome-wide identification and expression analysis of calcium-dependent protein kinase in maize. *BMC Genomics* **14**:433 DOI 10.1186/1471-2164-14-433.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35(6)**:1547–1549 DOI 10.1093/molbev/msy096.

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van Y, Rouzé P, Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research* **30(1)**:325–327 DOI 10.1093/nar/30.1.325.

Li AL, Zhu YF, Tan XM, Wang X, Wei B, Guo HZ, Zhang ZL, Chen XB, Zhao GY, Kong XY, Jia JZ, Mao L. 2008. Evolutionary and functional study of the *CDPK* gene family in wheat (*Triticum aestivum* L.). *Plant Molecular Biology* **66(4)**:429–443 DOI 10.1007/s11103-007-9281-5.

Liu W, Li W, He Q, Daud MK, Chen J, Zhu S. 2014. Genome-wide survey and expression analysis of calcium-dependent protein kinase in *Gossypium raimondii. PLOS ONE* **9(6)**:e98189 DOI 10.1371/journal.pone.0098189.

Loutfy N, ElTayeb MA, Hassanen AM, Moustafa MFM, Sakuma Y, Inouhe M. 2012. Changes in the water status and osmotic solute contents in response to drought and salicylic acid treatments in four different cultivars of wheat (Triticum aestivum). *Journal of Plant Research* **12(1)**:173–184 DOI 10.1007/s10265-011-0419-9.

Lu SX, Hrabak EM. 2013. The myristoylated amino-terminus of an *Arabidopsis* calcium-dependent protein kinase mediates plasma membrane localization. *Plant Molecular Biology* **82(3)**:267–278 DOI 10.1007/s11103-013-0061-0.

Martín ML, Busconi L. 2000. Membrane localization of a rice calcium-dependent protein kinase (CDPK) is mediated by myristoylation and palmitoylation. *The Plant Journal* **24(4)**:429–435 DOI 10.1046/j.1365-313x.2000.00889.x.

Meng L, Zhang Q, Yang J, Xie G, Liu JH. 2020. PtrCDPK10 of poncirus trifoliata functions in dehydration and drought tolerance by reducing ROS accumulation via phosphorylating PtrAPX. *Plant Science* **291**:110320 DOI 10.1016/j.plantsci.2019.110320.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29(9)**:e45 DOI 10.1093/nar/29.9.e45.

Ranty B, Aldon D, Cotelle V, Galaud JP, Thuleau P, Mazars C. 2016. Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. *Frontiers in Plant Science* **7**:327 DOI 10.3389/fpls.2016.00327.

Ray S, Agarwal P, Arora R, Kapoor S, Tyagi AK. 2007. Expression analysis of calcium-dependent protein kinase gene family during reproductive development and abiotic stress conditions in rice (*Oryza sativa* L. Ssp. indica). *Molecular Genetics and Genomics* **278(5)**:493–505 DOI 10.1007/s00438-007-0267-4.

Sallam A, Alqudah AM, Dawood MFA, Baenziger PS, Börner A. 2019. Drought stress tolerance in wheat and barley: advances in physiology, breeding and genetics research. *International Journal of Molecular Sciences* **20(13)**:3137 DOI 10.3390/ijms20133137.

Senthil-Kumar M, Mysore KS. 2014. Tobacco rattle virus-based virus-induced gene silencing in *Nicotiana benthamiana*. *Nature Protocols* **9(7)**:1549–1562 DOI 10.1038/nprot.2014.092.

Shi S, Li S, Asim M, Mao J, Xu D, Ullah Z, Liu G, Wang Q, Liu H. 2018. The *Arabidopsis* calcium-dependent protein kinases (CDPKs) and their roles in plant growth regulation and abiotic stress responses. *International Journal of Molecular Sciences* **19(7)**:1900 DOI 10.3390/ijms19071900.

Tanaka N, Fujita M, Handa H, Murayama S, Uemura M, Kawamura Y, Mitsui T, Mikami S, Tozawa Y, Yoshinaga T, Komatsu S. 2004. Proteomics of the rice-cell:systematic identification of the protein populations in sub-cellular compartments. *Molecular Genetics and Genomics* **271(5)**:566–576.

Thornton JW, DeSalle R. 2000. Gene family evolution and homology: genomics meets phylogen-etics. *Annual Review of Genomics and Human Genetics* **1**:41–73 DOI 10.1146/annurev.genom.1.1.41.

Vivek PJ, Resmi MS, Sreekumar S, Sivakumar KC, Tuteja N, Soniya EV. 2017. Calcium-dependent protein kinase in ginger binds with importin-α through its junction domain for nuclear localization, and further interacts with NAC transcription factor. *Frontiers in Plant Science* **7**:1909 DOI 10.3389/fpls.2016.01909.

Wang CT, Shao JM. 2013. Characterization of the *ZmCPK1* gene encoding a Calcium-dependent protein kinase responsive to multiple abiotic stresses in maize. *Plant Molecular Biology Reporter* **31**:222–230 DOI 10.1007/s11105-012-0496-5.

Wei S, Hu W, Deng X, Zhang Y, Liu X, Zhao X, Luo Q, Jin Z, Li Y, Zhou S, Sun T, Wang L, Yang G, He G. 2014. A rice calcium-dependent protein kinase *OsCPK9* positively regulates drought stress tolerance and spikelet fertility. *BMC Plant Biology* **14**:133 DOI 10.1186/1471-2229-14-133.

Yang Q, Rao J, Yi S, Meng K, Wu J, Hou Y. 2012. Antioxidant enzyme activity and chilling injury during low-temperature storage of Kiwifruit cv. Hongyang exposed

to gradual postharvest cooling. *Horticulture, Environment, and Biotechnology* **53**(**6**):505–512 DOI 10.1007/s13580-012-0101-8.

Yu J, Hui Y, Chen J, Yu H, Gao X, Zhang Z, Li Q, Zhu S, Zhao T. 2021. Whole-genome resequencing of 240 *Gossypium barbadense* accessions reveals genetic variation and genes associated with fiber strength and lint percentage. *Theoretical and Applied Genetics* **134**(**10**):3249–3261 DOI 10.1007/s00122-021-03889-w.

Yu TF, Zhao WY, Fu JD, Liu YW, Chen M, Zhou YB, Ma YZ, Xu ZS, Xi YJ. 2018. Genome-wide analysis of CDPK family in foxtail millet and determination of SiCDPK24 functions in drought stress. *Frontiers in Plant Science* **9**:651 DOI 10.3389/fpls.2018.00651.

Zhou L, Fu Y, Yang Z. 2009. A genome-wide functional characterization of *Arabidopsis* regulatory calcium sensors in pollen tubes. *Journal of Integrative Plant Biology* **51**(**8**):751–761 DOI 10.1111/j.1744-7909.2009.00847.x.

Zhu T, Liang C, Meng Z, Sun G, Meng Z, Guo S, Zhang R. 2017. CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biology* **17**(**1**):101 DOI 10.1186/s12870-017-1039-x.

Zhu SY, Yu XC, Wang XJ, Zhao R, Li Y, Fan RC, Shang Y, Du SY, Wang XF, Wu FQ, Xu YH, Zhang XY, Zhang DP. 2007. Two calcium-dependent protein kinases, CPK4 and *CPK11*, regulate abscisic acid signal transduction in *Arabidopsis*. *Plant Cell* **19**(**10**):3019–3036 DOI 10.1105/tpc.107.050666.

Zou JJ, Li XD, Ratnasekera D, Wang C, Liu WX, Song LF, Zhang WZ, Wu WH. 2015. Arabidopsis CALCIUM-DEPENDENT PROTEIN KINASE8 and CATALASE3 function in abscisic acid-mediated signaling and H_2O_2 homeostasis in stomatal guard cells under drought stress. *Plant Cell* **27**(**5**):1445–1460 DOI 10.1105/tpc.15.00144.

Zou JJ, Wei FJ, Wang C, Wu JJ, Ratnasekera D, Liu WX, Wu WH. 2010. Arabidopsis calcium-dependent protein kinase *CPK10* functions in abscisic acid and Ca^{2+}-mediated stomatal regulation in response to drought stress. *Plant Physiology* **154**(**3**):1232–1243 DOI 10.1104/pp.110.157545.