

Article

Genome wide identification of cotton C-repeat binding factor (CBF) and overexpression of *Gthu17439* (*GthCBF4*) gene confer cold stress tolerance in *Arabidopsis thaliana*

Jiangna Liu^{1,2}, Richard Odongo Magwanga^{1,3}, Yanchao Xu¹, Tingting Wei¹, Jie Zheng¹, Yuqing Hou¹, Yuhong Wang¹, Stephen Gaya Agong², Kunbo Wang¹, Zhongli Zhou^{1,*}, Xiaoyan Cai^{1,*}, and Fang Liu^{1,2,*}

¹ Zhengzhou Research Base, State Key Laboratory of Cotton Biology, Zhengzhou University/ Institute of Cotton Research, Chinese Academy of Agricultural Science, Anyang, China. JL (m18331098920@163.com); ROM (magwangarichard@yahoo.com); YX (xuyanchao2016@163.com); TTW (3091283731@qq.com); JZ(zhengjieself@163.com); YH (houyuqing18@163.com); YW (wangyh2525377@163.com); SGA (sgagong@jooust.ac.ke), KW (wkbcricri@163.com); ZZ (zhonglizhou@163.com); XC (cxycricri@163.com); FL (liufcricri@163.com)

² School of Agricultural Sciences, Zhengzhou University, Zhengzhou, China

³ School of Biological, Physical, Mathematics and Actuarial Sciences (SBPMAS), Main Campus, Jaramogi Oginga Odinga University of Science and Technology (JOUST), Main Campus, P.O. Box 210-40601 Bondo, Kenya

* Correspondence should be addressed to Fang Liu (liufcricri@163.com), Xiaoyan Cai (cxycricri@163.com) and Zhongli Zhou (zhonglizhou@163.com)

Abstract: Low temperature is a common biological abiotic stress in major cotton growing areas. Cold stress significantly affects the growth, yield and yield quality of cotton. Therefore, it is important to develop a more robust and cold stress tolerant cotton germplasm. Climate change and erratic climatic condition, plants have evolved various survival mechanisms, one of which induction of various stress responsive transcriptome factors, such as the C-repeat binding factor *GthCBF4*, which have been found to enhance cold tolerance in various plants. In this study detailed evaluation of the cotton C-repeat binding factor has been carried out. A total of 29, 28, 25, 21, 30, 26 and 15 proteins encoded by the C-repeat binding factor were identified in *G. herbaceum*, *G. arboreum*, *G. thurberi*, *G. raimondii*, *G. turneri*, *G. longicalyx* and *G. australe*, respectively. Phylogeny evaluation revealed that the proteins were grouped into seven clades, with clade 1 and 6 being the largest. Moreover, majority of the proteins encoded by the genes were predicted to be located within the nucleus, while some are distributed in other parts of the cell. Based on the transcriptome and RT-qPCR analysis, *Gthu17439* (*GthCBF4*) was highly upregulated and was further validated through forward genetics. The *Gthu17439* (*GthCBF4*) overexpressed plants showed a significantly tolerance to cold stress, with higher growth vigour compared to the wild types. The results showed that the *Gthu17439* (*GthCBF4*) could be playing a significant role in enhancing cold stress tolerance in cotton and can be further exploited in developing a more cold stress tolerance cotton germplasm

Keywords: CBF4; Transcription factors; Cold tolerance; overexpression; cotton

1. Introduction

Cotton is a thermophilic crop and is more sensitive to low temperatures [1]; China being the major cotton growing countries globally, the site specific regions within China, such as Xinjiang, is often affected by low temperature which significantly results in negative effects on plant growth and development [2]. Cold stress lead to inhibition of seed germination, reduction of plant growth and reproduction, as well as a decrease in crop yield and quality [3]. However, many crops, such as rice (*Oryza sativa*), maize (*Zea mays*), tomato (*Solanum lycopersicum*), soybean (*Glycine max*) and cotton (*Gossypium hirsutum*), do lack the ability to adapt to low temperature environments and can only grow

in tropical or subtropical regions [4–6]. In order to deter the emergence of the adverse effects, plants have evolved complex mechanisms to resist cold stress. by integrating the physiological, biochemical and molecular regulated mechanisms related to stress perception, signal transduction, gene expression and metabolic modifications to prevent cell oxidative damage [7]. In order to understand the effect of cold stress on plants, it is important to distinguish between cold (0-15°C) and freezing/chilling stress (< 0 °C). Cold stress leads to metabolic injury, destruction of the stability of protein complexes, inhibition of the metabolic pathways and various cellular processes in varying degrees, and eventually damage to the photosynthetic processes [8–10]. Freezing/chilling temperature promotes the ice formation in intercellular spaces of plant tissues [11]. When ice crystals are deposited on the cell wall, the extracellular water potential decreases and the cell membrane is destroyed, resulting in serious cell dehydration [12]. The most common phenomenon to resist the adverse effects of low temperature is adaptation. Many plants have increased freezing resistance after a period of low non-freezing temperature, a phenomenon known as cold acclimation [13–15]. Cold acclimation is an effective way to increase plant freezing resistance after a period of low non-freezing temperature [16]. When plants are exposed to chilling conditions, the plants do mobilize the cold-response genes (COR), which then activates.

C-repeat binding factors (CBF), followed by the accumulation of cryoprotectants, which results in the acquisition of freezing tolerance [7]. When plants are exposed to nonfreezing low temperature, CBF is rapidly induced by low temperature, and then the downstream target gene *COR* is activated [17]. In addition, as the expression of CBF is regulated by light quality, biological clock and photoperiod, it is necessary to understand the daily and seasonal regulation of CBF [18].

The C-repeat binding factor/dehydration-responsive element binding factor 1 (CBF/DREB1) is well-studied as a cold regulatory pathway. It is an adaptive response where plants increase freezing tolerance after exposure to low non-freezing temperatures [19]. The CBF/DREB protein, a DNA binding protein belonging to AP2/ERF superfamily, was first identified in *A. thaliana* as a LT-induced transcription factor [20]. APETALA2/ethylene-responsive element binding factor (AP2/ERF) family is a large class of plant-specific transcription factors, including AP2, RAV, ERF and CBF (C-repeat binding factor) [21,22]. The expression of *CBF* gene was very low at room temperature, but would increase rapidly within 15 minutes after plants were exposed to low temperature stimulation [4,16]. During cold acclimation, C-repeat binding factor (CBF) activated cold response (*COR*) gene and then accumulated cryoprotectants, which resulted in cold tolerance [7]. The *CBF / DREB* gene is activated by *CBF* expression (ICE) inducer through specific binding of cis elements to MYC in the promoter [16,20,23]. The cold reaction pathway of ICE-CBF-COR has been proved to be the effective defense mechanism to cold stress [4].

Six *CBF* family members have been identified in *A. thaliana*, *CBF1/DREB1C*, *CBF2/DREB1B*, and *CBF3/DREB1A* are induced by cold stress, while *CBF4/DREB1D*, *DREB1E/DDF2*, and *DREB1F/DDF1* are induced by osmotic stresses [24]. Moreover, the *CBF1*, *CBF2* and *CBF3* are arranged in tandem in an 8.7-kb region of the short arm of chromosome 4 in Arabidopsis [20]. All the three *CBF* Transcription factors perhaps acts as the activator of the expression of downstream *CORS* genes, but with different functions [25]. The *CBF1/2/3* triple mutant Arabidopsis seedlings obtained by CRISPR/Cas9 technology are more sensitive to low temperature than the single mutant *CBF2* and *CBF3*, as well as the double mutant *CBF1/3* [26], which is an indication that *CBF* genes play significant role enhancing cold stress tolerance in plants. The *CBF4* is a unique family member which plays an important role on drought tolerance in *A. thaliana* [22], but has been closely related to cold stress tolerance in grape [25]. In addition, the expression of *CBF4* have demonstrated significantly different among five different poplar species under cold stress, indicating *CBF4* transcription factors could be responsible for the regulatory role in cold stress response [27].

The CBF Transcription factors have been widely identified and isolated from rice, tomato, *Brassica napus*, wheat, barley and maize, which show that the CBF family is large in scale and complex in structure [2,28]. It has been reported that *AtCBF* gene was over-expressed in *Brassica napus* transgenic plants, and improved the cold tolerance [2,29]. The combination of *AtDREB1A* gene with stress-induced *RD29A* promoter improved the tolerance in transgenic tobacco to drought and low temperature stress [30,31]. Through phylogenetic analysis of Arabidopsis *CBFs* and its orthologous genes in other plants, it was found that *CBFs* were highly conservative in phylogeny [16].

As an important oil and fiber crop, cotton has been planted in more than 70 countries and plays an important role in the global economy. However, cotton yield is often adversely affected by biotic and abiotic stresses. Therefore, studying the molecular adaptation mechanism of cotton stress resistance, improving the stress resistance of cotton is of great significance for improving cotton yield. 21 CBF genes have been cloned from *G. hirsutum*; they can be divided into *GhCBF I*, *GhCBF II*, *GhCBF III* and *GhCBF IV* [24]. It provides useful clues for understanding the cold tolerance mechanism in cotton. However, due to the limited genome sequence, the expression profile of CBF family and its phylogenetic relationship with other plant CBF members are still unclear. In order to better understand the function and evolutionary relationship of CBF gene family in cotton, we analyzed the structural variation and evolution pattern of CBF family based on the genome-wide data of several cotton species, and explored the molecular mechanism of cold adaptation formation in *G. thurberi*. This study provides some ideas and reference for further research on the molecular mechanism of CBF gene regulating cold adaptation in cotton. The study of digging cotton endogenous genes from wild cotton as cold-tolerance-related genes and transferring them to tetraploid cultivars is beneficial to improve or enhance the genetic characteristics of existing varieties.

2. Results

2.1. Identification of CBF family genes in the cotton genome

The availability of the whole sequences for the seven cotton species enabled us to identify the CBF proteins harbored in their genome. The Pfam domain PF00847 was used as the query to obtain the CBF proteins, and finally get 29 members of *G. herbaceum*, 28 members of *G. arboreum*, 25 members of *G. thurberi*, 21 members of *G. raimondii*, 30 members of *G. turneri*, *G. longicalyx* has 26 members and *G. australe* has 15 members. Three representative cotton species from these seven species were chosen for further detailed analysis: *G. herbaceum*, *G. thurberi* and *G. australe*. The CBF CDS length in *G. herbaceum* range from 306 bp to 1,230 bp. In *G. australe*, it is 429 bp to 1,077 bp. In the physiochemical properties analysis of the CBF proteins, the results show a great difference. For the CBF proteins obtained from the *G. herbaceum*, their molecular weights ranged from 11,241.88 Da to 39,020.73 Da; isoelectric value ranged from 5.26 to 10.27. In *G. thurberi* molecular weights ranged from 16,179.23 Da to 45,591.53 Da, the *pI* ranged from 5.11 to 10.83. Similar in *G. australe*, it ranged from 15,676.36 Da to 40,867.96 Da. The CBF genes differed substantially by the encoded protein size and its biophysical properties (Table S1).

2.2. Phylogenetic analysis of cotton CBF gene family

In order to determine the phylogenetic relationship of CBF proteins, we constructed a phylogenetic tree by MEGA7.0, using the Neighbor-joining (NJ) method with minimal evolution and maximum parsimony. The CBF proteins were clustered into 7 clades and designated as clade 1 to 7 (Fig. 1). Clade 1 contains 61 CBF protein sequences at most, while clade 3 contains only 7 CBF amino acid sequences. Consistent with previous classification, all of the Arabidopsis CBFs were distributed among the clade 6 [2,22,29]. Except *G. australe* did not appear in clade 3, the other cotton species were distributed in all 7 groups.

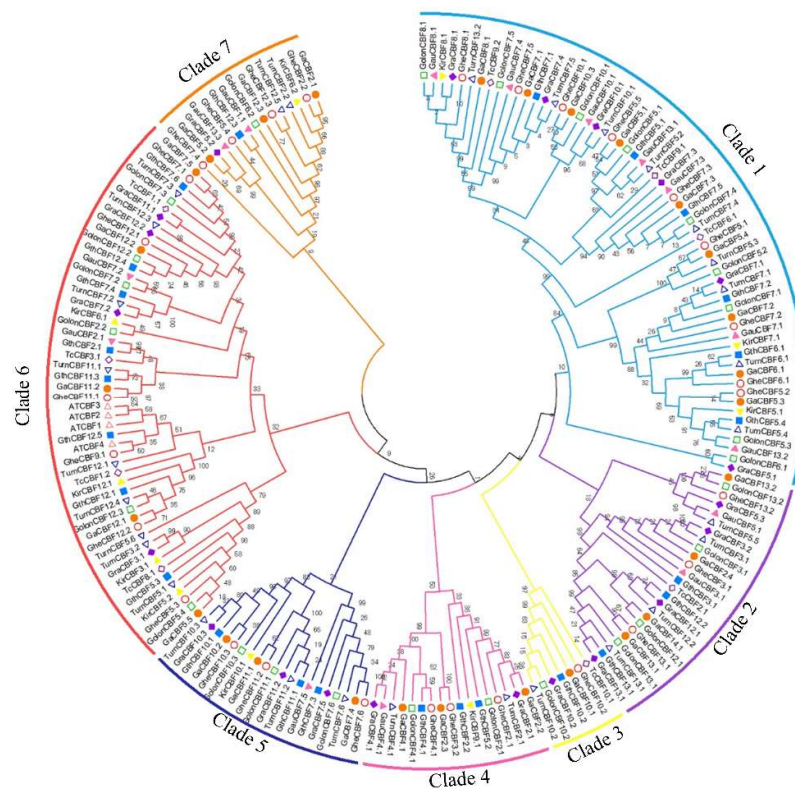


Fig. 1. Phylogenetic tree analysis of CBL transcription factors. Protein sequences alignment was done by ClustalX, the phylogenetic tree built by MEG7.0 software using 1000 bootstrap replication via Neighbor-joining approach. Different color was used to differentiate the homologous clusters.

2.3 Chromosomal mapping, Gene structure and C-terminal conserved motifs analysis

All the genes located on various chromosomes in the three cotton genomes, A, D and AD, and were named according to their position on the chromosome. In the A genome, *G. herbaceum* the CBF proteins were detected on 12 chromosomes, only Chr01 without CBF members. The most gene loci were detected in chromosome Chr05 and Chr07 (Fig. 2A). In the D genome, chromosome Gthu_1, Gthu_8 and Gthu_9 harbored no genes, Gthu_5, Gthu_7, and Gthu_12 had more gene loci (Fig 2B). Finally, in *G. australe* of the G genome, no CBF genes were found in chromosome G6, G9, G11 and G12, but the highest gene loci was only observed in chromosome G7 with 5 genes, respectively (Fig. 2C).

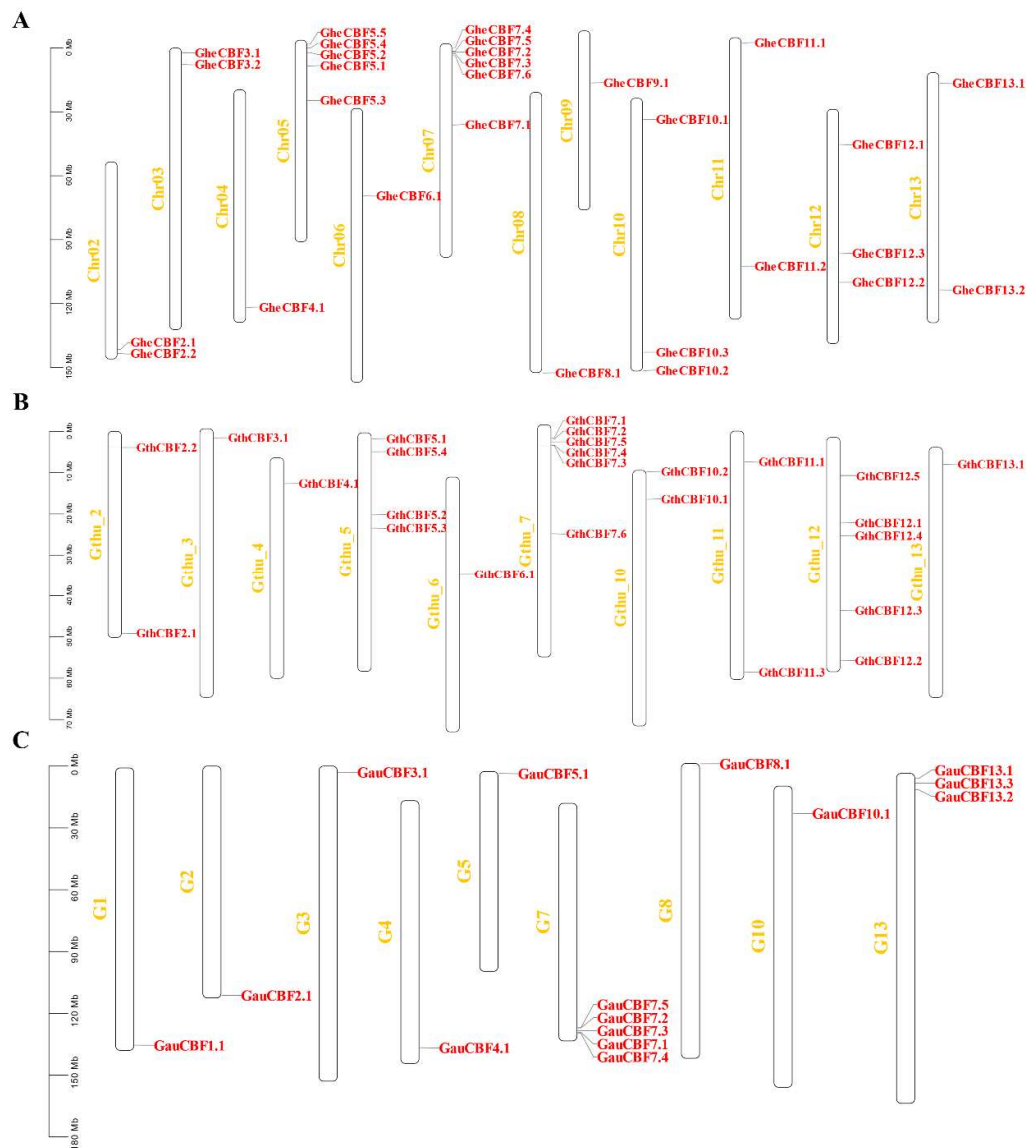


Fig 2: Chromosomal positions of the *CBF* genes. (A). Chromosomes of *G. herbaceum*, (B). Chromosomes of *G. thurberi* (C). Chromosomes of *G. australe*

We employed MEME to detect conserved motifs in the *CBF* family. There are 10 conserved motifs distributed in each *CBF* family (Fig. 3). Almost all *CBF* proteins have the same 3 motifs, motifs 1, 2, and 4. Analyzing the arrangement of exons and introns can provide important insights into the evolution of gene families [9]. To study the exon/intron structure of *CBF* gene, the CDS and the genome sequence were compared. The results showed that most of *G. herbaceum* contains one exon, *G. australe* only has one exon, and *G. thurberi* either contains 2 or more exons (Fig. 4A-D). Compared with phylogenetic analysis, most members in the same group have similar exon-intron structure and gene length.

In addition, through the online software Wolfspport (<https://wolfpsort.hgc.jp/>) we tried to determine the subcellular localization of the proteins encoded by the *CBF* genes. Among the three cotton species, the highest proportion of *CBF* protein is embedded in the nucleus. The locate signals of all the *CBF* members in *G. herbaceum* and *G. australe* are predicted in the nucleus but in *G. thurberi*, only some are in the nucleus. Except located in

the nucleus, the distribution of other CBF family members is very different. The chloroplast, mitochondria, plasma membrane, vacuole membrane and chloroplast also have the locate signals. (Table S2)

To study the exon/intron structure of *CBF* gene, the CDS and the genome sequence were compared. The results showed that most of *G. herbaceum* contains one exon, *G. australe* only has one exon, and *G. thurberi* either Contains 2 or more exons (Fig. 4A-D). Compared with phylogenetic analysis, most members in the same group have similar exon-intron structure and gene length.



Fig. 3. Conserved domain and gene structure analysis of CBF family protein sequences from the evolutionary level in cotton.

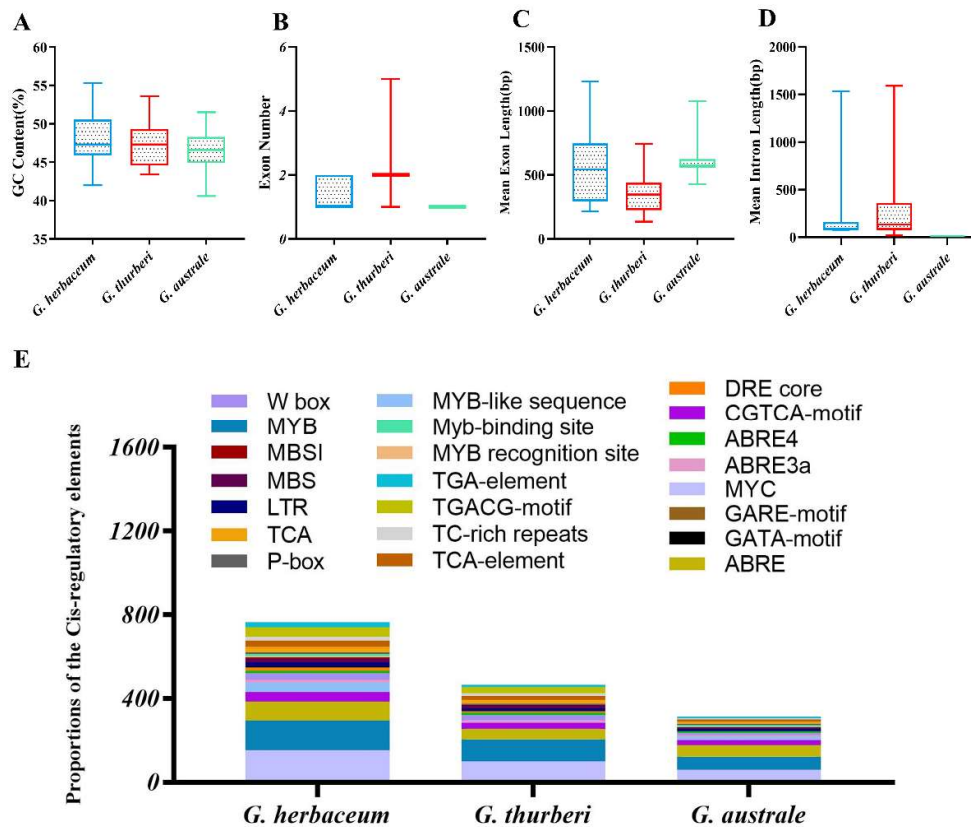


Fig. 4. Physiochemical properties and cis-regulatory element analysis of the proteins encoded by the cotton CBF genes. (A-D). GC content, exon number, mean exon length and mean intron length. (E). Cis-regulatory elements obtained for the various proteins encoded by the IQD genes in the *G. herbaceum*, *G. thurberi* and *G. australe*.

2.4 RNA-seq analysis and RT-qPCR validation of the CBF genes under cold stress conditions

G. thurberi transcriptome data was used to analyze the expression patterns of 24 CBF genes under cold stress (Fig. 5A). Among them, only 17 genes are differentially expressed under cold stress. According to *G. thurberi* transcriptome data, 12 differentially expressed CBF genes were selected, and 15 and 13 genes were selected from the two cotton varieties *G. herbaceum* and *G. australe*. RT-qPCR was detected to analyze the expression pattern; most genes are up-regulated in the three cotton species (Fig. 5B). Among the 12 genes in *G. thurberi*, there are 10 genes whose expression levels are up-regulated, while only 2 genes are down-regulated, but the trend is generally consistent with the transcriptome data. In *G. herbaceum*, 9 genes were up-regulated, and 6 were down-regulated. In *G. australe*, 8 were up-regulated and 5 were down-regulated. Integrate the transcriptome data and RT-qPCR result; we selected a highly expressed gene *GthCBF12.5* (*GthCBF4*) for further functional verification.

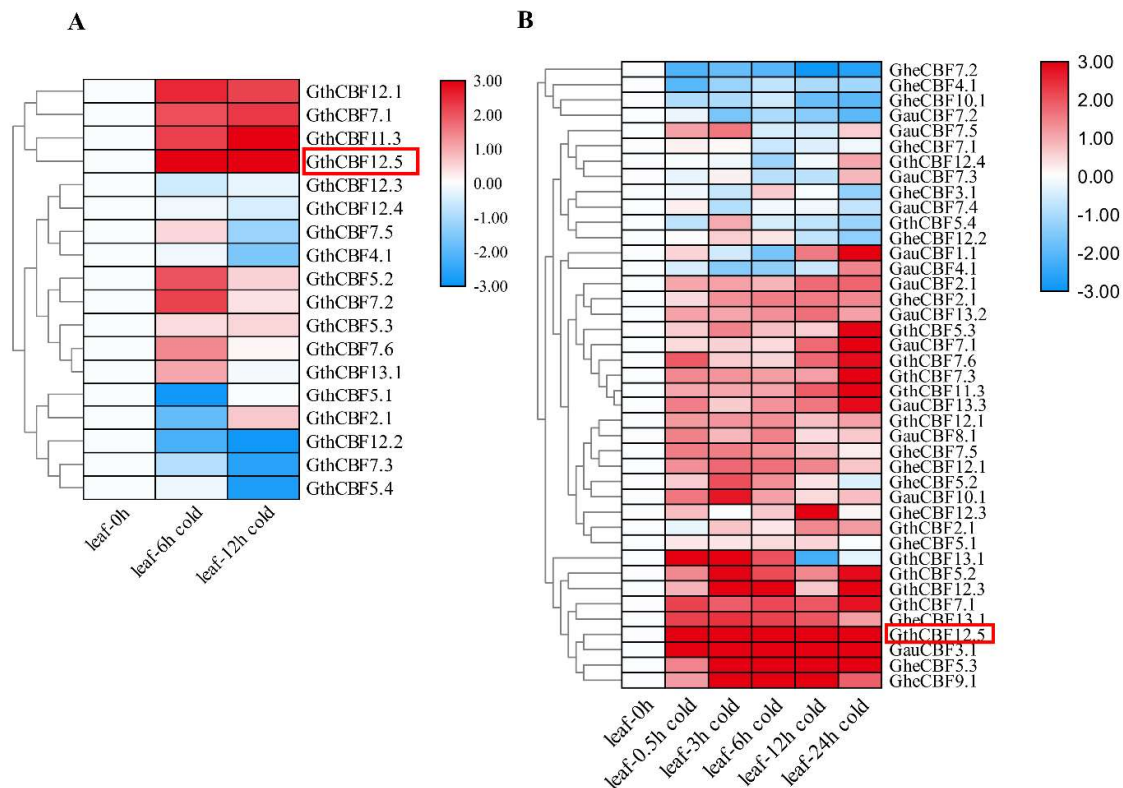


Fig. 5. RT-qPCR validations of transcript levels evaluated by RNA-Seq in *G. thurberi* under cold stress. (A) Heat map depicting log₂ (fold change) of transcription factor. (B) Heat map log₂ (fold change) of the CBF genes expression under cold stress condition. The samples were collected at 0h, 0.5h, 3h, 6h, 12h and 24 h for leaf tissues. Red: upregulated genes; blue: downregulated; Black: none expressed genes.

2.5 Analysis of subcellular localization results

The results showed that the control group showed green fluorescence signals on both the nucleus and cell membrane, while the fusion protein pCAMBIA2300-eGFP-Flag-*GthCBF4* only had green fluorescence signals in the nucleus (Fig. 6), indicating that the protein encoded by the gene was localized in the nucleus.

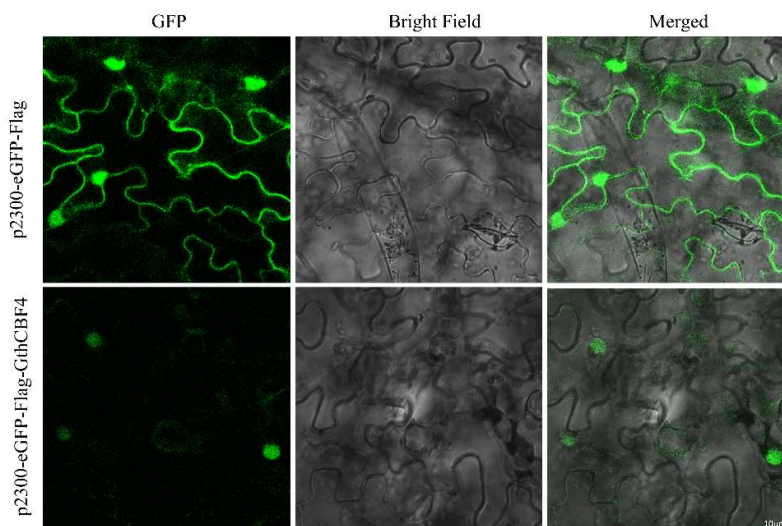


Fig. 6. Subcellular localization of pCAMBIA2300-eGFP-Flag-*GthCBF4* fusion protein

2.6. Phenotype identification and cell damage identification of overexpressed *Arabidopsis* under low temperature.

Two higher expression lines OE-1 and OE-3 were selected for phenotypic identification (Fig. 7A). Two-week-old transgenic and wild-type *Arabidopsis* plants were selected for cold tolerance test. As Fig.7B shown, two overexpression lines returned to green under the restoration culture, with new leaves growing, while most of the wild-type had withered and turned yellow. the survival rates of the two transgenic overexpression lines reached 60% and 63%, while the wild-type was only 8% (Fig. 7C). The expression analysis of *GthCBF4* was detected by real-time quantitative RT-qPCR after low temperature (-15 °C) treatment for 0, 1, 3 hours (Fig. 7D). The gene *GthCBF4* expression was significantly up-regulated in the overexpression lines, and the expression increased with the extension of treatment time. Trypan blue staining method was detected to reflect the cell damage under cold stress (Fig. 7E). Under normal conditions, the stained blue area on the leaves of the transgenic overexpression plants and wild-type plants was very small. While under cold treatment, the blue areas on transgenic leaves were significantly smaller than the wild-type, the color depth was also lighter.

DAB staining method was detected to reflect the accumulation of H₂O₂ in *Arabidopsis* leaves. the accumulation of H₂O₂ in the transgenic overexpression leaves and the wild-type were very low under normal conditions, and the production of brown matter was hardly seen. But after cold treatment, the brown area on the transgenic *Arabidopsis* leaves was obviously larger than the wild type, and the color depth was also deeper (Fig. 7E).

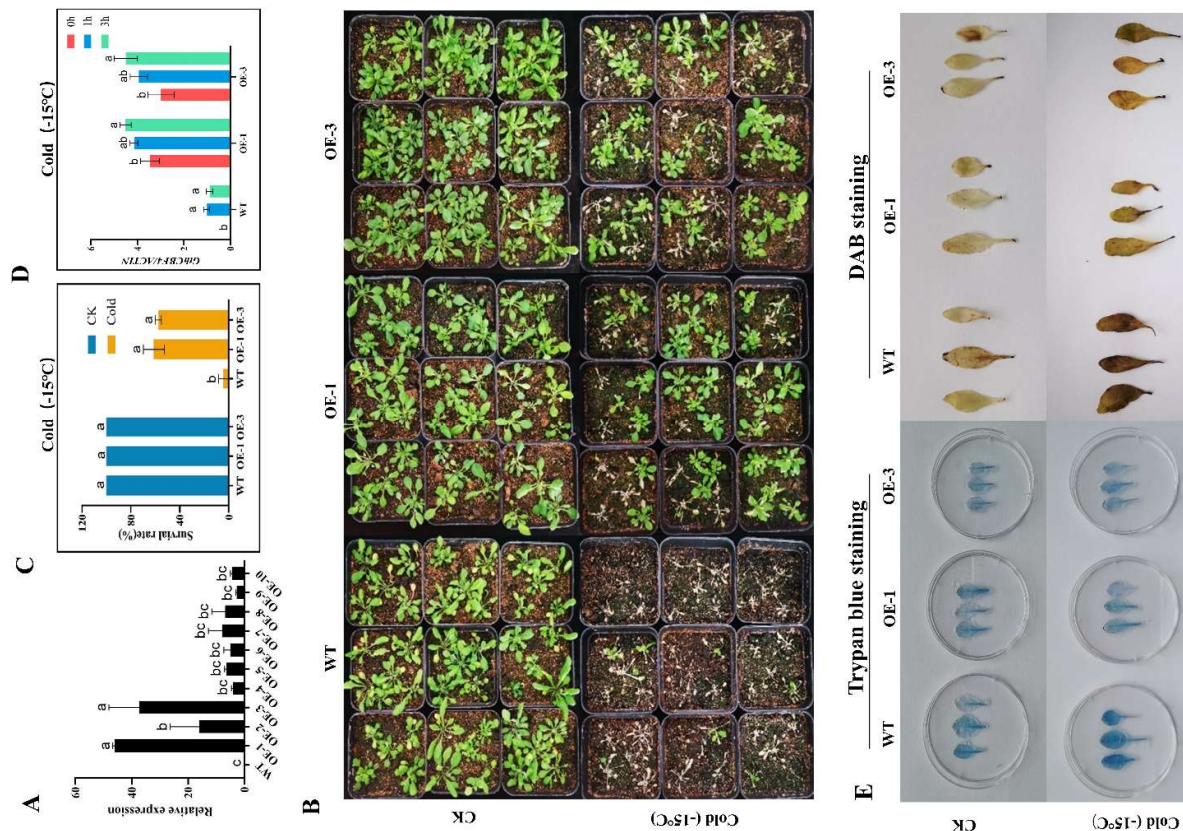


Fig. 7. Phenotype identification and cell damage identification of overexpressed *Arabidopsis* under low temperature. (A) Quantitative expression level of *GthCBF4* gene in transgenic *Arabidopsis*. (B) -15°C treatment restores the growth status of wild-type and transgenic *Arabidopsis thaliana* after culture. (C) Freezing survival assay. (D) log10 (fold change) of the *GthCBF4* genes expression in -15°C treatment. (E) Trypan blue staining and DAB staining.

2.7 Overexpression plants and evaluation of physio-morphological traits under low temperature environment

As Fig. 8 A and B showed that no wild type lines could germinate after treatment, while the germination rates of the two overexpression lines are 16% and 7%, respectively. On the other hand, the root length of the two overexpression lines is significantly longer than the WT (Fig. 8C, D). We further evaluated known stress responsive genes such as the *COR15A*, *RD29A*, *KIN1* and *COR47* [26]. After treatment wild type and *CBF* triple mutants with low-temperature (-15°C) for 0h, 1h and 3h, the expression of *COR* genes was assessed by real-time quantitative RT-qPCR. All the four abiotic stress responsive genes were significantly upregulated in the overexpression plants under low-temperature (-15°C) stress conditions (Fig. 8E).

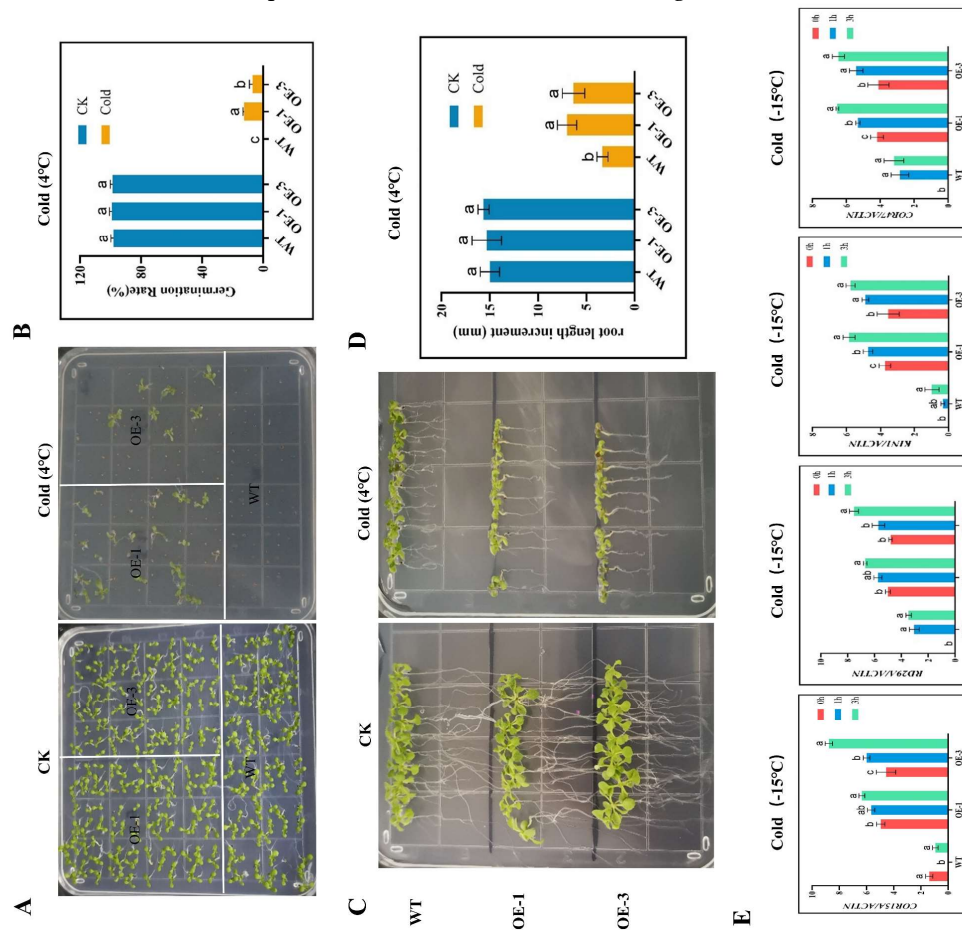


Fig. 8. Determination of the growth phenotypes and regulatory gene expression of transgenic Arabidopsis under cold stress conditions. (A-B) Germination rate determination. (C-D) Root length determination. (E) The expression levels of abiotic stress responsive gene.

3. Discussion

Cotton is an important economic crop, but the production of cotton is affected by various abiotic stress factors, especially the yield under drought, salinity and cold stress. Although abiotic stress is a major challenge in cotton growth, there is no detailed study on *CBF* gene in cotton [32]. In previous studies, *CBF* family have been identified in cotton (*Gossypium hirsutum*) [24], wheat [33], lettuce [34], *Brassica napus* [35], Barley [36] and soybean [37], but there is few study in diploid cotton. In this work, genome wide identification, characterization and functional analysis of the proteins encoded by the cotton *CBF* genes was done, identified in the three cotton species, with 29, 25 and 15 *CBF* pro-

teins in *G. herbaceum*, *G. thurberi* and *G. australe*, respectively. The results obtained showed that the proteins encoded by the *CBF* genes in cotton were higher compared to other plants such as lettuce with 14, brassica napus with 10 and soybean with 14 *CBF* genes, but less than hexaploid wheat with 65 *CBF* genes and barley with 20 *CBF* genes. Studies have found that *Arabidopsis* *CBF* family members have an AP2 domain, and each has a conserved amino acid sequence upstream and downstream of AP2. The upstream of AP2 is PKK/PKKPAGR (RAGRxxKFx ETRHP) and the downstream is DSAWR [38]. If the PKK/PKKPAGR mutation can inhibit the binding ability of *CBF* to the *COR* promoter of its downstream genes, thereby weakening the level of *CBF* regulation, this sequence is necessary for *CBF* to perform its transcription factor function. Cotton genome contains a large and complex *CBF* subfamily, they contain conserved AP2/EREBP domains and *CBF* characteristics, indicating that cotton *CBFs* have similar function with other *CBFs* in dicotyledons [25]. Phylogenetic analysis showed that the *CBF* family was divided into seven groups, among these genes *CBF1*, *CBF2* and *CBF3* are all induced by cold stress in *Arabidopsis thaliana*. Therefore, we speculate that *CBF* genes may also respond to abiotic stress, especially to cold. Further analysis showed that the isolated *CBF* gene was highly expressed under cold stress, consistent with previous research results.

In plants, the transcriptional regulation of osmotic stress response mainly depends on two main cis-regulatory elements, which are related to stress response genes ABREs and dehydration response elements (DREs). DREs are mainly involved in ABA-independent pathways, and ABRE is responsible for detecting ABA-mediated osmotic stress signals. In our work, we found that the *CBF* genes in the three cotton species are rich in stress response genes (ABRE), dehydration response element (DRE) and low temperature response element (LTR). It is speculated that exogenous environmental stress can induce the expression of *CBF* gene through its response to cis-acting elements, and further improve the resistance of plants to environmental stress. The results of sub-cellular localization confirmed that the fusion gene pCAMBIA2300-eGFP-Flag-*GthCBF4* was expressed in the nucleus of tobacco leaf epidermal cells, and the *GthCBF4* protein was localized in the nucleus, which confirmed previous studies and was consistent with the results of bioinformatics analysis.

Overexpression of *Arabidopsis* *CBF* gene in other plant species or overexpression of *CBF* of other species in *Arabidopsis* has revealed the potential of the *CBF* genes in enhancing frost resistance [39]. Moreover, it shows *CBF* gene plays an important role in plant cold tolerance. Recently, the expression of *CBF1* and *CBF3* were down regulated with the RNAi and antisense technique, resulting in a 25-50% reduction in cold-treated plants [24]. In this study, the *GthCBF4* was strongly up-regulated in cotton seedlings under low temperature treatment. The survival rate of *GthCBF4* transgenic *Arabidopsis thaliana* plants was significantly improved after freezing treatment. From the trypan blue staining and DAB staining, it can be seen that the staining of overexpression plants should be lighter; indicating that overexpression of the *GthCBF4* gene can reduce the damage to plants caused by freezing treatment. It strongly proves *GthCBF4* is the key gene related to cold stress. According to the measurement results of the germination rate and root elongation, it can be seen that under the same cold treatment, the two overexpression lines and the wild type are significantly different, which further proves that the gene *GthCBF4* can enhance the ability to resist cold damage. In generally, the expression level of these four *COR* genes is correlated with the freezing tolerance level. Proof of previous research *COR15A*, which encodes a chloroplast-targeted polypeptide, strengthens the cold resistance of chloroplasts [40]. Previous studies have confirmed that the *CBF* gene can induce the expression of *COR47*, *RD29A* and *KIN1* genes to improve the cold tolerance of plants [41]. It can be seen that the expression levels of stress response genes in the two overexpression lines are significantly higher than the wild type. And with the extension of the freezing treatment time, the gene expression level is also increasing. The up-regulation of the stress responsive genes further confirmed that the overexpression plants ability to tolerate the effects of cold was significantly strength. In

the past two decades, the transcriptional network of CBF signaling pathway has been extensively studied. Although many *COR* genes have been identified in genome-wide expression profiles, only about 10-25% of them are regulated by CBF, which means that more early cold regulated transcription factors are involved in improving cold tolerance. More research is needed to identify more transcription regulators and explore their potential relationships to increase our understanding of cold related transcription networks.

4. Materials and Methods

4.1 Identification of CBF family genes in the cotton genome

Wild diploid cotton species data was downloaded from CottonGen (<https://www.cottongen.org/>) to construct a local BLAST database. The Arabidopsis CBF protein sequences were used as probes to compare with the wild diploid cotton. The E-value threshold for BLASTP was set at $1e^{-10}$ to obtain the final dataset of CBF proteins. Then, the Pfam (<http://pfam.sanger.ac.uk/search>) and SMART (<http://smart.embl-heidelberg.de/>) databases were detected to confirm each predicted CBF protein sequence [42]. Redundant sequences and incomplete sequences were removed. The sequences of 10 *Gossypioides kirkii* and 9 *Theobroma cacao* CBF proteins were obtained from the CottonGen (<https://www.cottongen.org/>) databases, respectively. In addition, physicochemical parameters including the molecular weight (MW) and isoelectric point (*pI*) of each gene product were calculated using compute the *pI/Mw* tool from ExPASy (<http://www.expasy.org/tools/>).

4.2 Sequence alignment and phylogenetic analysis of cotton CBF gene family

An alignment of multiple CBF protein sequences from *A. thaliana*, *Gossypioides kirkii* and *Theobroma cacao* was generated using the ClustalW program. A neighbor-joining analysis of the generated alignment was performed using the unweighted pair-group method with arithmetic mean algorithm to construct an unrooted phylogenetic tree. Bootstrap value was 1000, and other parameters were used by default value. The tree was visualized with MEGA 7.0 software [43].

4.3 Gene structure and C-terminal conserved motifs analysis

Structural information for the *CBF* genes includes chromosomal location and gene length. Exons and introns were predicted by comparing the coding sequences with genomic sequences. The conserved motif analysis of CBF protein sequence was predicted by MEME online software. Use the CDD (<https://www.ncbi.nlm.nih.gov/cdd/>) database to search for the conserved domain information of CBF, and use the TBtools mapping tool to draw the conserved domains [44].

4.4 Retrieval and analysis of promoter sequences

The 2000 bp sequence upstream of ATG were extracted the transcription start site of the *CBF* gene sequence, and submit the obtained sequence to the PlantCARE website (<http://www.dna.affrc.go.jp/PLACE/signalscan.html>). Identification of possible cis-acting elements in the promoter region is used to identify putative cis-regulatory elements in the promoter sequence [45]. In addition, we carried out the subcellular localization prediction of all the CBF proteins by an online tool WoLFPSORT (<http://www.genscript.com/wolf-psort.html>) [46].

4.5 RNA extraction and qRT-PCR analysis

The total RNA was extracted using EASYspin plus plant RNA kit (Aidlab, Biotech, Beijing, China), following the manufacturer's instructions. The Nanodrop 1000spectrophotometer was used to determine the quantity and quality of the RNA samples. The primers used for qRT-PCR were designed using primer premier 5 software for all genes using Primer Premier 5 Software (Table S3). The cotton *GhActin* gene, forward primer sequence 5'ATCCTCCGCTCTTGACCTTG3' and reverse primer sequence 5'TGTCCGTCAGGCAACTCAT3'), was used as a reference gene for the analysis. Real-time PCR reactions were carried out in a final volume of 25 μ l, using a SYBR Green

master mix and an ABI7500 thermal cycler (Applied Biosystems, Foster City, CA, U.S.A.), following manufacturer's instructions. The analysis of each sample was done; three technical replicates and biological replicates of each sample were taken for the analysis. Each sample was analyzed in triplicate. Three biological replications of the experiment were carried out.

4.5.1 Plant material

The seeds of *G. herbaceum*, *G. thurberi* and *G. austral* (Depilation and cut outting seed shell as pretreat) germinated in sand at 25°C for 4 days. Then seedlings were transferred to the hydroponic facility equipped with Hoagland nutrient solution [47]. The greenhouse conditions were 28°C during the day/25°C at night, the photoperiod was 16 hours, and the relative humidity was 60-70%. In the three-leaf stage, cotton seedlings were kept at 4°C under normal light, and then the leaves were harvested under cold treatments of 0, 0.5, 3, 6, 12 and 24 h. Each treatment was repeated three times. The leaf samples were immediately collected in liquid nitrogen, frozen and stored at -80°C until RNA extraction.

4.5.2 *GthCBF4* subcellular localization analysis

To explore the subcellular localization of the *GthCBF4* gene, a pCAM-BIA2300-eGFP-Flag-*GthCBF4* fusion vector of CBF and GFP was constructed and transiently expressed in the epidermal cells of tobacco leaves, which was driven by the 35s promoter and transformed *Agrobacterium* LBA4404 competent cells. In addition, *Agrobacterium* competent cells expressing only the *GFP* gene were used as a negative control. Four-week-old tobaccos cotyledon flat leaves were selected for infusion, and cultivated in dark for 24h-36h after infiltration, and fluorescence observation was performed under a laser confocal microscope.

4.6 Functional verification of *GthCBF4* in *Arabidopsis* Screening of Transgenic *Arabidopsis*

To explore the function of *GthCBF4* gene, this study constructed a pBI121-*GthCBF4* recombinant vector and transformed into the promoter cells of *A. tumefaciens* GV3101 by adopting a freeze-thaw method [48,49]. The wild-type *Arabidopsis thaliana* was soaked by the dipping method, and used 50 mg/mL kanamycin for positive selection until the T3 generation. In the T2 generation seedlings, the expression lines of gene *GthCBF4* were screened by RT-qPCR, and two high-expressing transgenic lines were obtained.

4.7 Phenotypic identification, physiological and biochemical parameters of WT and overexpressed plants under low temperature stress

The transgenic and wild-type *Arabidopsis thaliana* grown in plastic bowls for two weeks and their growth is basically the same are selected for treatment. The *Arabidopsis thaliana* was placed in an environment of -15°C for 3 hours. After the treatment, the treated plants were moved to a 4°C light incubator to thaw for 4 hours, and finally they were cultivated under normal light conditions at 23°C. After 7 days, take pictures and count the survival rate of the plants. For each tissue, at least three independent biological replicates were carried out. The germination rate and root length was determined under cold stress; a t-test was used to verify the significance of the difference in root length between the mutant and the wild types. The trypan blue staining method was adopted to reflect the cell damage of transgenic *Arabidopsis* under cold stress. Moreover, the DAB staining was applied to reflect the accumulation of peroxidase in transgenic *Arabidopsis* under cold stress. DAB staining was performed using DAB chromogenic kit (Nanjing Jiancheng Bioengineering Institute, Nanjing, China).

5. Conclusions

In conclusion, through the family analysis of diploid wild cotton CBF, we determined the distribution of CBF proteins with 29 in *G. herbaceum*, 25 in *G. thurberi* and 15 in *G. australe*. According to the results of RNA-seq and RT-QPCR, a highly up-regulated gene *GthCBF4* under cold stress was selected. Combined with the functional verification

test of overexpression of *Arabidopsis thaliana*, it is proved that *GthCBF4* plays an important role in improving the cold resistance of *G. thurberi*. This result provides a solid foundation for further research on the molecular function of *GthCBF4* protein in cotton, and provides ideas for further improving the role of *CBF* gene in the cold stress regulatory network.

Supplementary Materials: Table S1. Physiochemical properties of the *CBL* genes. Physiochemical properties and cis-regulatory element analysis of the proteins encoded by the cotton *CBF* genes. (A-D). GC content, exon number, mean exon length and mean intron length. (E). Cis-regulatory elements obtained for the various proteins encoded by the *IQD* genes in the *G. herbaceum*, *G. thurberi* and *G. australe*.; Table S2. Subcellular localization of the proteins encoded by the *CBL* genes; Table S3: List of primers for RT-qPCR analysis

Author Contributions: Conceptualization, JL, ROM. and FL.; methodology, JL,ROM, YX,TTW and FL software, JL,ROM, YX,TTW and FL.; validation, JL,ROM, YX,TTW and FL. and Z.Z.; formal analysis, JL,ROM, YX,TTW and FL.; investigation, JL,ROM, YX,TTW and FL.; resources, SAG, KW, ZZ, XC and FL.; data curation, JL,ROM, YX,TTW and FL.; writing—original draft preparation, JL,ROM, YX,TTW and FL.; writing—review and editing, JL,ROM, YX,TTW, JZ, YH, SGA, KW, ZZ, XC and FL.; visualization, JL,ROM, YX,TTW and FL.; supervision, SGA, KW, ZZ, XC and FL.; project administration, SGA, KW, ZZ, XC and FL.; funding acquisition, SGA, KW, ZZ, XC and FL. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (32072023), Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences

Acknowledgments: We are deeply indebted to the staff and students of the germplasm resources for their help in the course of the research work in addition to the data and analysis process.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hemantaranjan, A.; Bhanu, A. N.; Singh, M. N.; Yadav, D. K.; Patel, P. K.; Singh, R.; Katiyar, D. Heat stress responses and thermotolerance. *2014*, *1*, 62–70, doi:10.15406/apar.2014.01.00012.
- Rihan, H. Z.; Al-issawi, M.; Fuller, M. P. Advances in physiological and molecular aspects of plant cold tolerance. *J. Plant Interact.* **2017**, *12*, 143–157, doi:10.1080/17429145.2017.1308568.
- Körner, C. Plant adaptation to cold climates [version 1 ; referees: 2 approved] Referee Status: **2016**, *5*, 1–5, doi:10.12688/f1000research.9107.1.
- Wang, D.-Z.; Jin, Y.-N.; Ding, X.-H.; Wang, W.-J.; Zhai, S.-S.; Bai, L.-P.; Guo, Z.-F. Gene regulation and signal transduction in the ICE-CBF-COR signaling pathway during cold stress in plants. *Biochem.* **2017**, *82*, 1103–1117, doi:10.1134/S0006297917100030.
- Xiong, L.; Schumaker, K. S.; Zhu, J. Cell Signaling during Cold , Drought , and Salt Stress. *Plant Cell* **2002**, *supplement*, S165–S184, doi:10.1105/tpc.000596.S166.
- Kumar Verma, J.; Wardhan, V.; Singh, D.; Chakraborty, S.; Chakraborty, N. Genome-wide identification of the alba gene family in plants and stress-responsive expression of the rice alba genes. *Genes (Basel)*. **2018**, *9*, doi:10.3390/genes9040183.
- Thomashow, M. F. Plant cold acclimation: Freezing tolerance genes and regulatory mechanisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 571–599, doi:10.1146/annurev.arplant.50.1.571.
- Bruce, T. J. A.; Matthes, M. C.; Napier, J. A.; Pickett, J. A. Stressful “memories” of plants: Evidence and possible mechanisms. *Plant Sci.* **2007**, *173*, 603–608, doi:10.1016/j.plantsci.2007.09.002.
- Colebrook, E. H.; Thomas, S. G.; Phillips, A. L.; Hedden, P. The role of gibberellin signalling in plant responses to abiotic stress. *J. Exp. Biol.* **2014**, *217*, 67–75, doi:10.1242/jeb.089938.
- Zhao, Q.; Xiang, X.; Liu, D.; Yang, A.; Wang, Y. Tobacco Transcription Factor NtHHLH123 Confers Tolerance to Cold Stress

- by Regulating the NtCBF Pathway and Reactive Oxygen Species Homeostasis. *Front. Plant Sci.* **2018**, *9*, 1–13, doi:10.3389/fpls.2018.00381.
11. Sciences, N. Plant Freezing and Damage. **2001**, doi:10.1006/anbo.2000.1352.
 12. Uemura, M. Plant plasma membrane proteomics for improving cold tolerance. **2014**, doi:10.3389/fpls.2013.00090.
 13. Hussain, H. A.; Hussain, S.; Khaliq, A.; Ashraf, U.; Anjum, S. A.; Men, S.; Wang, L. Chilling and Drought Stresses in Crop Plants: Implications, Cross Talk, and Potential Management Opportunities. *Front. Plant Sci.* **2018**, *9*, 1–21, doi:10.3389/fpls.2018.00393.
 14. Kumar, S.; Nayyar, H.; Bhanwara, R. K.; Upadhyaya, H. D. Chilling stress effects on reproductive biology of chickpea. *J. SAT Agric. Res.* **2010**, *8*, 1–14.
 15. Gene, O. Improved Abiotic Stress Tolerance in Plants By Accumulation of. *Plant Biotechnol.* **2005**.
 16. Shi, Y.; Ding, Y.; Yang, S. Molecular Regulation of CBF Signaling in Cold Acclimation. *Trends Plant Sci.* **2018**, *23*, 623–637.
 17. Zhang, Z.; Huang, R. Enhanced tolerance to freezing in tobacco and tomato overexpressing transcription factor TERF2/LeERF2 is modulated by ethylene biosynthesis. *Plant Mol. Biol.* **2010**, *73*, 241–249, doi:10.1007/s11103-010-9609-4.
 18. Gong, Z.; Xiong, L.; Shi, H.; Yang, S.; Xu, G.; Chao, D.; Li, J.; Wang, P.; Herrera-estrella, L. R. Plant abiotic stress response and nutrient use efficiency. **2020**, *63*, 635–674.
 19. Liu, Q.; Kasuga, M.; Sakuma, Y.; Abe, H.; Miura, S.; Yamaguchi-shinozaki, K. Two Transcription Factors , DREB1 and DREB2 , with an EREBP / AP2 DNA Binding Domain Separate Two Cellular Signal Transduction Pathways in Drought- and Low- Temperature-Responsive Gene Expression , Respectively , in Arabidopsis. **1998**, *10*, 1391–1406.
 20. Jin, Y.; Zhai, S.; Wang, W.; Ding, X.; Guo, Z.; Bai, L.; Wang, S. Identification of genes from the ICE–CBF–COR pathway under cold stress in Aegilops–Triticum composite group and the evolution analysis with those from Triticeae. *Physiol. Mol. Biol. Plants* **2018**, *24*, 211–229, doi:10.1007/s12298-017-0495-y.
 21. Jin, J.-H.; Wang, M.; Zhang, H.-X.; Khan, A.; Wei, A.-M.; Luo, D.-X.; Gong, Z.-H. Genome-wide identification of the AP2/ERF transcription factor family in pepper (*Capsicum annuum* L.). *Genome* **2018**, *61*, doi:10.1139/gen-2018-0036.
 22. Haake, V. Transcription Factor CBF4 Is a Regulator of Drought Adaptation in Arabidopsis. *PLANT Physiol.* **2002**, *130*, 639–648, doi:10.1104/pp.006478.
 23. Chinnusamy, V.; Ohta, M.; Kanrar, S.; Lee, B. ha; Hong, X.; Agarwal, M.; Zhu, J. K. ICE1: A regulator of cold-induced transcriptome and freezing tolerance in arabidopsis. *Genes Dev.* **2003**, *17*, 1043–1054, doi:10.1101/gad.1077503.
 24. Ma, L. F.; Zhang, J. M.; Huang, G. Q.; Li, Y.; Li, X. B.; Zheng, Y. Molecular characterization of cotton C-repeat/dehydration-responsive element binding factor genes that are involved in response to cold stress. *Mol. Biol. Rep.* **2014**, *41*, 4369–4379, doi:10.1007/s11033-014-3308-1.
 25. Siddiqua, M.; Nassuth, A. Vitis CBF1 and Vitis CBF4 differ in their effect on Arabidopsis abiotic stress tolerance , development and gene expression. **2011**, *1*, 1345–1359, doi:10.1111/j.1365-3040.2011.02334.x.
 26. Zhao, C.; Zhang, Z.; Xie, S.; Si, T.; Li, Y.; Zhu, J. Mutational Evidence for the Critical Role of CBF Transcription Factors in Cold Acclimation in Arabidopsis 1. **2016**, *171*, 2744–2759, doi:10.1104/pp.16.00533.
 27. Genetics, M.; Li, Y.; Xu, B.; Du, Q.; Zhang, D. under abiotic stress Association genetics and expression patterns of a CBF4 homolog in Populus under abiotic stress. **2014**, doi:10.1007/s00438-014-0967-5.
 28. Yu, X.; Pijut, P. M.; Byrne, S.; Asp, T.; Bai, G.; Jiang, Y. Plant Science Candidate gene association mapping for winter survival and spring regrowth in perennial ryegrass. *Plant Sci.* **2015**, *235*, 37–45, doi:10.1016/j.plantsci.2015.03.003.
 29. Bai, G. H. S. M. L. Signal transduction during cold , salt , and drought stresses in plants. **2011**, doi:10.1007/s11033-011-0823-1.
 30. Bhatnagar-Mathur, P.; Vadez, V.; Sharma, K. K. Transgenic approaches for abiotic stress tolerance in plants: Retrospect and prospects. *Plant Cell Rep.* **2008**, *27*, 411–424, doi:10.1007/s00299-007-0474-9.

31. Lata, C.; Yadav, A.; Pras, M. Role of Plant Transcription Factors in Abiotic Stress Tolerance. *Abiotic Stress Response Plants - Physiol. Biochem. Genet. Perspect.* **2011**, doi:10.5772/23172.
32. Bo, H.; Longguo, J. I. N.; Jinyuan, L. I. U. Molecular cloning and functional characterization of a DREB1 / CBF-like gene (GhDREB1L) from cotton. **2007**, *50*, 7–14, doi:10.1007/s11427-007-0010-8.
33. Mohseni, S.; Che, H.; Djillali, Z.; Dumont, E.; Nankeu, J.; Danyluk, J. Wheat CBF gene family: identification of polymorphisms in the CBF coding sequence. **2012**, *881*, 865–881, doi:10.1139/gen-2012-0112.
34. Sunchung, P.; Ainong, S.; Mo, B. Genome-wide identification and expression analysis of the CBF / DREB1 gene family in lettuce. **2020**, 1–14, doi:10.1038/s41598-020-62458-1.
35. Lv, K.; Li, J.; Zhao, K.; Chen, S.; Zhang, W.; Liu, G.; Wei, H. Plant Science Overexpression of an AP2 / ERF family gene , BpERF13 , in birch enhances cold tolerance through upregulating CBF genes and mitigating reactive oxygen species. **2020**, *292*, doi:10.1016/j.plantsci.2019.110375.
36. Marquez-cedillo, L.; Filichkin, T.; Amundsen, K.; Stockinger, E. J.; Thomashow, M. F.; Chen, T. H. H.; Hayes, P. M. Structural , functional , and phylogenetic characterization of a large CBF gene family in barley. **2005**, 533–551, doi:10.1007/s11103-005-2498-2.
37. Kidokoro, S.; Watanabe, K.; Ohori, T.; Moriwaki, T.; Maruyama, K.; Mizoi, J.; Myint Phyu Sin Htwe, N.; Fujita, Y.; Sekita, S.; Shinozaki, K.; Yamaguchi-Shinozaki, K. Soybean DREB1/CBF-type transcription factors function in heat and drought as well as cold stress-responsive gene expression. *Plant J.* **2015**, *81*, 505–518, doi:10.1111/tpj.12746.
38. Ofuku, K. D. I. J. The AP2 domain of APETALA 2 defines a large new family of DNA binding proteins in Arabidopsis. **1997**, *94*, 7076–7081.
39. Tondelli, A.; Francia, E.; Barabaschi, D.; Pasquariello, M.; Pecchioni, N. Plant Science Inside the CBF locus in Poaceae. *Plant Sci.* **2011**, *180*, 39–45, doi:10.1016/j.plantsci.2010.08.012.
40. Thalhammer, A.; Bryant, G.; Sulpice, R.; Hinch, D. K.; Planck, M.; Plant, M.; Potsdam, D.; Germany, A. T. Disordered Cold Regulated15 Proteins Protect Chloroplast Membranes during Freezing through Binding and Folding , But Do Not Stabilize Chloroplast Enzymes in Vivo 1 [W][OPEN]. **2014**, *166*, 190–201, doi:10.1104/pp.114.245399.
41. Soliman, M. H.; Alayafi, A. A. M.; Kelish, A. A. El; Elsaoud, A. M. A. Acetylsalicylic acid enhance tolerance of Phaseolus vulgaris L . to chilling stress , improving photosynthesis , antioxidants and expression of cold stress responsive genes. *Bot. Stud.* **2018**, 1–17, doi:10.1186/s40529-018-0222-1.
42. Letunic, I.; Doerks, T.; Bork, P. SMART: recent updates , new developments and status in 2015. **2015**, *43*, 257–260, doi:10.1093/nar/gku949.
43. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874, doi:10.1093/molbev/msw054.
44. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H. R.; Frank, M. H.; He, Y.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, doi:10.1016/j.molp.2020.06.009.
45. Higo, K.; Ugawa, Y.; Iwamoto, M.; Korenaga, T. Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res.* **1999**, *27*, 297–300.
46. Horton, P.; Park, K. J.; Obayashi, T.; Fujita, N.; Harada, H.; Adams-Collier, C. J.; Nakai, K. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **2007**, *35*, doi:10.1093/nar/gkm259.
47. Hoagland, D. R.; Aron, D. I. *The water-culture method for growing plants without soil*; 1950; Vol. 347; ISBN 347.
48. Corbin, C.; Lafontaine, F.; Sepúlveda, L. J.; Carqueijeiro, I.; Courtois, M.; Lanoue, A.; Dugé de Bernonville, T.; Besseau, S.; Glévarec, G.; Papon, N.; Atehortúa, L.; Giglioli-Guivarc'h, N.; Clastre, M.; St-Pierre, B.; Oudin, A.; Courdavault, V. Virus-induced gene silencing in Rauwolfia species. *Protoplasma* **2017**, *254*, 1813–1818, doi:10.1007/s00709-017-1079-y.
49. Corbin, C.; Lafontaine, F.; Sepúlveda, L. J.; Carqueijeiro, I.; Courtois, M.; Lanoue, A.; Dugé de Bernonville, T.; Besseau, S.;

Glévarec, G.; Papon, N.; Atehortúa, L.; Giglioli-Guivarc'h, N.; Clastre, M.; St-Pierre, B.; Oudin, A.; Courdavault, V.
Virus-induced gene silencing in Rauwolfia species. *Protoplasma* **2017**, *254*, 1813–1818, doi:10.1007/s00709-017-1079-y.