

# Characterization of the Gh4CL gene family reveals a role of Gh4CL7 in drought tolerance

**Shichao Sun**

Shihezi University College of Agriculture

**Xianpeng Xiong**

Shihezi University College of Agriculture

**Xiaoli Zhang**

Shihezi University College of Agriculture

**Hongjie Feng**

Chinese Academy of Agricultural Sciences Cotton Research Institute

**Qianhao Zhu**

CSIRO Agriculture and Food

**Jie Sun** (✉ [sunjie@shzu.edu.cn](mailto:sunjie@shzu.edu.cn))

Shihezi University <https://orcid.org/0000-0002-8068-8102>

**Yanjuan Li**

Shihezi University College of Agriculture

---

## Research article

**Keywords:** Gossypium hirsutum, 4CL, Transgenic Arabidopsis, Drought stress, VIGS

**Posted Date:** November 21st, 2019

**DOI:** <https://doi.org/10.21203/rs.2.17480/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Plant Biology on March 23rd, 2020.  
See the published version at <https://doi.org/10.1186/s12870-020-2329-2>.

# Abstract

**Background** The function of 4-coumarate-CoA ligases (4CL) under abiotic stresses have been studied in plants, however, limited is known about the upland cotton 4CL genes and their roles in response to drought stress.

**Results** In this study, we performed genome-wide identification of the 4CL genes in *G. hirsutum* and investigated the expression profiles of the identified genes in various cotton tissues and in response to stress conditions with an aim to identify 4CL gene(s) associated with drought tolerance. We identified 34 putative 4CL genes in *G. hirsutum* that were clustered into three classes. Genes of the same class usually share a similar gene structure and motif composition. Many cis -elements related to stress and phytohormone responses were found in the promoters of the Gh4CL genes. Of the 34 Gh4CL genes, 26 were induced by at least one abiotic stress and 10 were up-regulated under the PEG simulated drought stress conditions, including Gh4CL7 . Virus-induced gene silencing (VIGS) in cotton and overexpression (OE) in *Arabidopsis thaliana* were applied to investigate the biological function of Gh4CL7 in drought tolerance. The Gh4CL7 -silencing cotton plants showed more sensitive to drought stress, probably due to decreased relative water content (RWC), chlorophyll content and antioxidative enzyme activity, increased stomatal aperture, and the contents of malondialdehyde (MDA) and hydrogen peroxide ( $H_2O_2$ ). *Arabidopsis* lines overexpressing Gh4CL7 , however, were more tolerant to drought treatment, which was associated with improved antioxidative enzyme activity, reduced accumulation of MDA and  $H_2O_2$  and up-regulated stress-related genes under the drought stress conditions. In addition, compared to their respective controls, the Gh4CL7 -silencing cotton plants and the Gh4CL7- overexpressing *Arabidopsis* lines had a lower and higher lignin content, respectively. Taken together, these results demonstrated that Gh4CL7 could positively respond to drought stress and therefore might be a candidate gene for improvement of drought tolerance in cotton.

**Conclusion** We comprehensively identified 4CL family genes in upland cotton and provide valuable information on the evolutionary relationship. Gh4CL7 gene is involved in lignin biosynthesis and play a positive role in relieving drought stress.

## Background

Cotton is an important cash crop in many developing countries and frequently grown in dry lands or on supplementary irrigation [1], because agricultural water consumption can no longer be expanded thanks to water competition among domestic, industrial and agricultural users. The quantity and quality of fiber produced by cotton plants are directly related to water available to them during their developmental stages. When suffered from water deficits, especially during the period of flowering and fructification, cotton would show significant loss in yield, sometimes up to 50% reduction compared to those that have been irrigated [2, 3].

In the long-term evolutionary history, plants have formed a complex gene-metabolic network to accommodate a variety of environmental changes. As an important metabolite, lignin plays vital roles in defense against biotic and abiotic stresses [4–7]. Lignin is synthesized through the phenylpropane pathway. The 4-coumarate-CoA ligase (EC 6.2.1.12, 4CL) is the main branch point enzyme of the phenylpropanoid pathway, which catalyzes cinnamic acid to generate corresponding CoA thioesters [8]. Products of 4CL subsequently serve as substrates of various oxygenases, reductases and transferases for biosynthesis of lignin, flavonoids, anthocyanins, aurones, stilbenes, coumarins, suberin, cutin, sporopollenin, and others [9, 10]. The *4CL* gene family has been characterized in many plants, such as *Arabidopsis* [11], rice [12, 13], soybean [14] and aspen [15, 16]. Genes of the *4CL* family in dicots can be classified into two distinct groups, type I and type II [8, 17]. Type I genes are mainly involved in lignin biosynthesis whereas type II genes are involved in biosynthesis of phenylpropanoids other than lignin. Some additional genes containing the same conserved motifs of *4CLs* and showing high similarity with the 4CL proteins are classified as *4CL-like* genes [18].

Studies have shown that *4CL* genes play momentous roles in plants, such as regulation of growth and development, protection against biotic and abiotic stresses [12, 19, 20]. In *Arabidopsis*, *At4CL1*, *At4CL2*, and *At4CL4* were found to be involved in lignin formation, the *4cl1 4cl2* double and *4cl1 4cl2 4cl3* triple mutant plants exhibited a dwarf and bushy phenotype [21]. In rice, *Os4CL2* was specifically expressed in anthers and induced by UV irradiation [13]. *Plagiochasma appendiculatum* thallus plants showed downregulation of *Pa4CL1* when treated with abscisic acid (ABA), and showed upregulation of *Pa4CL1* when treated with salicylic acid and MeJA [22]. In both poplars and *Arabidopsis*, the expression levels of *4CL* genes were induced by salt stress and wounding [7, 8]. The *4CL-like* genes may also play a role in response to abiotic stresses [23–24]. Overexpression of *Fm4CL-like1* in tobacco increased drought tolerance due to increasing lignin accumulation and the activities of antioxidant enzymes, and upregulating the expression levels of stress-related genes [24]. Nevertheless, our knowledge of the *4CL* gene family in cotton is very limited.

To gain insights into the cotton *4CL* gene family and its role in abiotic stress tolerance, in this study, we did genome-wide identification of *4CL* genes in *G. hirsutum* and analysed their expression changes in response to various abiotic stresses based on publically available RNA-seq datasets. We identified 34 putative *Gh4CL* genes in *G. hirsutum* and 26 of them were found to be induced by at least one stress, including *Gh4CL7* that was up-regulated under polyethylene glycol (PEG) osmotic stress. We further investigated the function of *Gh4CL7* in drought tolerance by silencing its expression in cotton using virus-induced gene silencing (VIGS) and generating transgenic *Arabidopsis* plants overexpressing *Gh4CL7*. Our results indicated that *Gh4CL7* functions positively in response to drought stress and is a potential candidate gene for improving drought resistance of cotton by genetic engineering.

## Results

# Genome-wide identification and phylogenetic analysis of *Gh4CL* genes

Using the approach described in Materials and Methods, we identified 34 *Gh4CL* genes in *G. hirsutum*. They are distributed on 22 chromosomes and an unanchored scaffold that was not assigned to a particular chromosome (Table 1). We named them *Gh4CL1* to *Gh4CL34* based on their chromosomal location. The protein length of Gh4CLs is between 276 and 615 amino acids (aa) with ORF from 828 to 1845 bp, molecular weight from 30 to 67.73 KD, and pI from 5.3 to 9.76 (Table 1). Based on subcellular localization prediction, 22 Gh4CL proteins are located in biomembrane, 11 in microbody (peroxisome), and the remaining one in cytoplasm. All the Gh4CL proteins have two structural domains, a putative AMP-binding domain “SSGTTGLPKG” (Box I) and a conserved domain “GEICIRG” (Box II) (Additional file 1: Fig. S1). Phylogenetic analysis using the 34 *Gh4CL* genes and *4CL* genes from *A. thaliana*, *G. max*, *P. tremuloides*, *P. trichocarpa*, *R. idaeus*, and *I. tinctoria* showed that they were clustered into three groups (Fig. 1a).

## Chromosomal distribution and syntenic relationship analysis

The 34 *Gh4CL* genes were found to be randomly distributed in *G. hirsutum* chromosomes (Fig. 1b), and there was no correlation between the length of chromosome and the number of *Gh4CL* genes. It has been suggested that both tandem duplication and segmental duplication events might have contributed to the generation of gene family with multiple members [25]. We found that two pairs of *Gh4CL* genes, *Gh4CL10/11* and *Gh4CL21/22*, were in tandem configuration on chromosomes A09 and D03, respectively. Segmental duplication could be involved in generation of 12 *Gh4CLs* based on MCScanX analysis. To better understand the evolutionary selection on *Gh4CL* genes, the Ka/Ks ratio was calculated for each homologous/paralogous gene pair (Additional file 2: Table S1). All gene pairs had a ratio of Ka/Ks < 1, suggesting that these *Gh4CL* genes have experienced purifying selective pressure during their evolution history to eliminate deleterious mutations.

## Analyses of gene structure and motifs

To provide greater insight into the gene structures of *Gh4CLs*, we analysed the intron–exon arrangement of the 34 *Gh4CL* genes based on their cDNA and the corresponding genomic sequences (Fig. 2a). Similar to the *4CL* genes in other plant species, all *Gh4CL* genes have a multi-exon and multi-intron structure, containing 3 to 11 exons. Of the 34 *Gh4CL* genes, 14 have six exons, and one each contains 3, and 11 exons, respectively. Genes in the same class have similar gene structure, especially in homologous/paralogous pairs.

All Gh4CL protein sequences were subjected to MEME analysis, and 20 distinct motifs were identified, they were designated motif 1 to 20 (Fig. 2b). Of the 20 motifs, 14 motifs encode an AMP-binding enzyme domain and six motifs were unknown (Additional file 2: Table S2). As shown in Fig. 2b, motifs 1–13, 17, 20 were found in all three classes, whereas other 5 motifs were found to be present in only one or two classes. For example, motif 14 was found only in class I, motifs 15 and 19 were found only in class II, and motif 18 was found only in class III. As expected, most of the structurally similar members within the same subfamily shared common motifs, indicating potential functional similarity between the paralogous Gh4CL pairs, for example, Gh4CL7 and Gh4CL25, and Gh4CL13 and Gh4CL31.

## Analysis of *cis*-elements in the promoters of *Gh4CL* genes

*Cis*-elements in combination with transcription factors regulate the transcription level of a gene. To identify potential *cis*-elements involved in regulation of transcription of *Gh4CL* genes, we scanned *cis*-elements in the promoter region (2 kb upstream of ATG) of each *Gh4CL* gene using the online tool PlantCARE [26]. The results were shown in Fig. 3. Many *Gh4CL* genes harbored plant hormone-responsive elements, including ABA responsive elements (ABREs), auxin responsive elements (AuxRR-core, TGA-elements and TGA-box), MeJA-responsive elements (CGTCA-motif, TGACG-motif), gibberellin-responsive elements (TATC-box, GARE-motif and P-box), and salicylic acid responsive elements (TCA-elements). Out of the 34 *Gh4CL* genes, 22 possessed MeJA-responsive elements, 15 contained gibberellin-responsive elements, 13 had ABA responsive element, 12 harbored salicylic acid responsive elements, and 7 genes possessed auxin responsive element. Moreover, low-temperature responsive elements (LTR), defense and stress responsiveness elements (TC-rich repeats) and drought-responsive elements (MBS) were also identified in the promoter regions of some *Gh4CL* genes. Out of the 34 *Gh4CL* genes, 15 contained TC-rich repeats, 11 had LER and 7 contained MBS, indicating that these genes may be involved in response to abiotic stresses. We were interested in the 7 genes containing MBS, including *Gh4CL7*, *Gh4CL8*, *Gh4CL13*, *Gh4CL23*, *Gh4CL25*, *Gh4CL26* and *Gh4CL34* because they may play a role in response to drought stress.

## Tissue specific expression patterns of *Gh4CL* genes

The expression patterns in various tissues provide clue for the possible biological functions of genes of interest. We thus analysed the transcript abundance of the *Gh4CL* genes in different tissues (root, stem, leaves, flower, ovule and fibers at 5, 10, 15 and 20 DPA) under normal growth conditions using the publically available RNA-seq data [27]. We found that 10 *Gh4CL* genes were expressed in all the tested tissues [base on fragments per kilobase of transcript per million mapped reads (FPKM)  $\geq 1$ ], and 4 genes (*Gh4CL3*, *Gh4CL5*, *Gh4CL18* and *Gh4CL27*) showed weak or no expression in all tissues analysed (Fig. 4). In addition, 8 *Gh4CL* genes (*Gh4CL2*, *Gh4CL4*, *Gh4CL8*, *Gh4CL12*, *Gh4CL17*, *Gh4CL24*, *Gh4CL29* and *Gh4CL30*) were highly expressed (FPKM  $\geq 20$ ) in stem, with the highest expression level observed for

*Gh4CL17* (FPKM  $\geq 84$ ) and 6 genes (*Gh4CL7–8*, *Gh4CL12*, *Gh4CL20* and *Gh4CL30–31*) were strongly expressed in leaves, with the highest expression observed for *Gh4CL20* (FPKM  $\geq 202$ ).

## Expression analysis of *Gh4CL* genes under different abiotic stress conditions

Since *4CL* genes are capable of responding to biotic and abiotic stresses in various plant species, we investigated the transcript abundance of the *Gh4CL* genes under cold, heat, polyethylene glycol (PEG) and salt stresses using the transcriptomic data of *G. hirsutum* [27]. We found that 26 *Gh4CL* genes were induced significantly by one or more stresses, and the remaining 8 *Gh4CL* genes (*Gh4CL3*, *Gh4CL5*, *Gh4CL10*, *Gh4CL18–19*, *Gh4CL23*, *Gh4CL28* and *Gh4CL34*) were not induced by either of the four stresses (Fig. 5a). Comparing the four stress conditions, more *Gh4CL* genes showed altered expression in response to salinity, cold and heat stresses than to PEG stress. Notably, ten *Gh4CL* genes (*Gh4CL2*, *Gh4CL7–9*, *Gh4CL11–13*, *Gh4CL17*, *Gh4CL22*, *Gh4CL25* and *Gh4CL31*) showed increased expression (treatment FPKM/control FPKM  $\geq 1.5$ ) in response to PEG stress over the 3 h to 12 h time period. To verify these results, we investigated the expression patterns of the selected *Gh4CL* genes under the simulated drought treatment using qRT-PCR. As shown in Fig. 5b, the expression levels of *Gh4CL7–8*, *Gh4CL12–13*, *Gh4CL17*, *Gh4CL22* and *Gh4CL24* were up-regulated in cotton leaves over the time period of 3h to 24h after PEG stress, consistent with the RNA-seq based results.

## Silencing of *Gh4CL7* compromises cotton's tolerance to drought stress

Based on the analysis results of promoter *cis*-elements and expression patterns under drought stress, three *Gh4CL* genes, including *Gh4CL7*, *Gh4CL8* and *Gh4CL13*, were considered as candidate genes involved in the regulation of drought stress response in cotton. To verify our hypothesis, *Gh4CL7* was selected for further functional analysis. We used VIGS to silence the expression of *Gh4CL7* using the TRV vectors (*TRV:Gh4CL7*) that was confirmed by qRT-PCR (Additional file 1: Fig. S2). *TRV:GhCHLI* was used as a positive control (Additional file 1: Fig. S3).

Phenotypic difference between the *TRV:Gh4CL7* and *TRV:00* plants was observed after 20 days of water deficiency treatment. Compared to the *TRV:00* plants, the *TRV:Gh4CL7* plants displayed severe wilting and yellowing leaves (Fig. 6a), consistent with a lower leaf RWC (Fig. 6b) and a decrease chlorophyll contents (Fig. 6c). Besides, it was also found that the size and the ratio of width to length of stomata significantly increased in the *TRV:Gh4CL7* plants (Fig. 6d-f), which might accelerate the transpiration rate under drought conditions, consistent with the observed higher WLR (Fig. 6g).  $H_2O_2$  content and MDA level were measured to reflect the cell damage or injury in *TRV:Gh4CL7* and *TRV:00* plants. During drought stress, the *TRV:Gh4CL7* plants accumulated more malondialdehyde (MDA; Fig. 6h) and hydrogen peroxide ( $H_2O_2$ ; Fig. 6i) compared to the *TRV:00* plants. SOD, POD and CAT activity in the *TRV:Gh4CL7*

and *TRV:00* plants were also measured to explore the function of *Gh4CL7* in the modulation of antioxidant enzymes (Fig. 6j). As expected, under drought stress conditions, the *TRV:Gh4CL7* plants displayed a significantly reduced SOD, POD and CAT activity as compared to the *TRV:00* plants. Additionally, six stress-related genes (*GhABI4*, *GhABF4*, *GhLEA14*, *GhRD22*, *GhRD29* and *GhNCED1*) were down-regulated in the *TRV:Gh4CL7* plants after drought treatment (Additional file 1: Fig. S4). These results suggested that silencing of *Gh4CL7* decreases cotton's tolerance to drought stress.

## Overexpression of *Gh4CL7* in *Arabidopsis* enhances drought tolerance

We further investigated the function of *Gh4CL7* in response to drought stress using *Arabidopsis* plants overexpressing *Gh4CL7*. Three independent *Gh4CL7-OE* lines that showed an elevated level of *Gh4CL7* (Fig. 7a) were selected for phenotyping under the drought stress conditions. Compared to the wild-type (WT) plants, the three *Gh4CL7-OE* lines had a decreased germination rate (Fig. 7b), but showed a significantly increased root length under the mannitol stress conditions (Fig. 7c, d). Three-weeks-old seedlings of *Gh4CL7-OE* and WT were used for water deficiency treatment. No obvious phenotypic difference was observed between *Gh4CL7-OE* and WT in the mock group. However, the *Gh4CL7-OE* plants showed much less damage than WT after 10 days of water deficiency (Fig. 7e). Transgenic *Arabidopsis* leaves had a slower WLR *in vitro* compared to the WT plants (Fig. 7f). Additionally, under drought stress conditions, the H<sub>2</sub>O<sub>2</sub> content and MDA level in the *Gh4CL7-OE* plants were relatively lower than that in WT (Fig. 7g-h), but the SOD, POD and CAT activities were significantly higher (Fig. 7i). To further elucidate the possible mechanism of *Gh4CL7* in response to drought stress, the transcript levels of four known ABA-responsive genes (*AtRD29B*, *AtRD22*, *AtABI4*, *AtCOR15A*) and two ABA-biosynthesis genes (*AtNCED3* and *AtNCED5*) were analyzed in the *Gh4CL7-OE* lines and WT plants after drought stress treatment. The qRT-PCR data showed that the expression levels of these genes were induced in *Gh4CL7-OE*, but not or just slightly induced in WT by drought stress (Additional file 1: Fig. S5). These results indicated that overexpression of *Gh4CL7* could enhance the tolerance of transgenic *Arabidopsis* plants to drought stress.

## *Gh4CL7* plays an important role in lignin biosynthesis

*Gh4CL7* belongs to class I whose genes have been shown to regulate lignin synthesis [21, 28]. We investigated whether or not *Gh4CL7* is also involved in lignin biosynthesis by comparing the lignin contents of the *Gh4CL7-OE Arabidopsis* lines and *TRV:Gh4CL7* cotton plants with that of their corresponding control plants. The lignin content increased by approximately 10% in the *Gh4CL7-OE* lines compared with WT plants (Fig. 8a), while decreased by approximately 20% in the *TRV:Gh4CL7* plants compared with the *TRV:00* plants (Fig. 8b). Additionally, the stem of the *TRV:Gh4CL7* plants were sectioned and stained with phloroglucinol-HCl to detect the presence of lignin (Fig. 8c). We found that the stem section of the *TRV:Gh4CL7* plants with reduced lignin content exhibited light red color, but the

*TRV:00* plants displayed typically purple-red color after phloroglucinol-HCl staining. These results suggested that *Gh4CL7* is related to lignin synthesis. We also analysed the expression level of other genes related to lignin synthesis by the phenylpropane pathway, including *GhPAL*, *GhCCoAOMT1*, *GhCOMT1*, *GhCOMT2*, *GhCOMT3*, *GhCCR1*, *GhCCR2*, and *GhCAD*. The relative expression level of these genes were lower in *TRV:Gh4CL7* cotton plants than in *TRV:00* (Fig. 8d), indicating that *Gh4CL7* could affect the accumulation of lignin by regulating the transcription level of these downstream genes of the lignin biosynthesis pathway.

## Discussion

The *4CL* gene family has been characterized in several plants, including *Arabidopsis thaliana*, *Populus trichocarpa*, *Oryza sativa* and *Glycine max* [12, 29–31]. Genes of this family have been reported to function not only in plant growth and development [13, 28, 32–33], but also in response to biotic and abiotic stresses [7, 34, 35]. However, no comprehensive analysis of the *4CL* genes has been documented in *G. hirsutum*. In this study, we did genome-wide identification of *4CL* genes in *G. hirsutum* and investigated their expression profiles in various tissues and under different stress conditions with an aim to identify *4CL* gene(s) with a potential role in stress tolerance. In total, 34 *Gh4CL* members were identified in the *G. hirsutum* genome (Table 1). In other plant species, such as *A. thaliana*, *4CL* genes were divided into class I, class II and class 4CL-like [18]. The 34 *Gh4CL* genes could also be clustered into three classes. We named the *4CL-like* class as class III, which contains the largest number of *Gh4CL* genes (in total 25) together with *At4CL6–9*, *At4CL11*, and *li4CL1* (Fig.1a). The class  $\emptyset$  *Gh4CLs* cannot catalyze any of the hydroxycinnamic acid substrates into the corresponding CoA esters, they may exhibited a function different from that of class I (relate to the lignin biosynthesis) and class II (relate to the biosynthesis of flavonoids) *4CL* genes [8, 11, 17, 18, 23]. Multiple sequence alignments revealed that all 4CL-like proteins contained similar structural components, e.g. conserved Box I and Box II domains without known specific biochemical function (Additional file 1: Fig. S1) [36, 37]. Gene organization plays an important role in the evolution of multiple gene families [38, 39]. Gene structure analysis showed that *Gh4CL* genes of the same class share a similar intron-exon structure (Fig. 2b), similar to the observations made in other gene families [40, 41].

*Cis*-elements located in the promoter region of genes play key roles in the developmental and environmental regulation of gene expression [42]. According to *cis*-element analysis, the promoter regions of *Gh4CL* genes possess elements related to stress responses, such as ABRE, TC-rich, LTR, MBS, TGA-element, TCA-element, CGTCA-motif and TGACG-motif [43–46], suggesting a potential role of the *Gh4CL* genes with these *cis*-element(s) in tolerance of stresses, such as drought, salt, heat, ABA, and low temperature. *4CL* genes have been shown to be involved in response to stresses in other plants [19, 34, 35]. Based on transcriptome data, 26 *Gh4CL* genes were differentially expressed between the stress treatment and the mock, and many *Gh4CL* genes were induced by drought, including *Gh4CL7* (Fig. 5a). The promoter of *Gh4CL7* contains an MBS *cis*-acting element, which may be associated with its up-regulation upon drought treatment (Fig. 5b). In addition, the *Gh4CL* genes showed different expression



profiles in different tissues (Fig. 4), suggesting that they perhaps play a broad range of roles in cotton growth and development.

Abiotic stresses often disrupt the balance between ROS production and clearance in cells, leading to increased ROS concentrations and oxidative damage to biofilms, proteins, DNA, and RNA, thereby inhibiting plant growth and development [47, 48]. Therefore, scavenging ROS is essential for plants to resist abiotic stresses.  $H_2O_2$  is one of the ROS, and its over accumulation in the plant cell can cause oxidative damage, while its lower concentration correlates with drought tolerance [49]. In the *Gh4CL7* gene-silenced plants, the  $H_2O_2$  content was found to be increased significantly under drought stress, so was the MDA level, an indicator of ROS destructive effects [50]. SOD, POD, and CAT are antioxidant enzymes in plant cells which scavenge the toxic ROS and lead to enhanced tolerance under stress conditions [51]. We found a lower SOD, POD, and CAT activity in the *TRV:Gh4CL7* plants than in the *TRV:00* plants indicating that the reduced ability of the *Gh4CL7* silenced plants to scavenge ROS that might have led to membrane damages and chlorophyll content reduction. On the other hand, the *Gh4CL7* overexpression plants had a lower level of  $H_2O_2$  and MDA and a higher activity of antioxidant enzymes (SOD, POD, and CAT) compared to WT under the drought stress conditions. These results are consistent with previous finding that *Fm4CL* had a role in drought tolerance by modulating the level of ROS [24].

Drought stress affects crop yield and quality through its negative influence on seed germination, seedling growth, photosynthesis, and transpiration [1]. We found that *A. thaliana* plants overexpressing *Gh4CL7* had longer roots but a lower germination rate than WT plants under the osmotic stress conditions, suggesting that *Gh4CL7* played a negative role in seed germination, but a positive role in promoting root elongation under the osmotic stress conditions. Breeding crops with thriving and deeper root systems is the goal of geneticists and breeders, because it can increase productivity of crops under drought conditions [52]. Longer roots might be a result of changed ABA signaling pathway, which play a crucial role in root development under drought stress [53–55], as well as many other factors related to drought responses, including stomata closure and stress-gene regulation [53, 56–58]. Under drought stress conditions, the closed stomata can decrease transpiration rate that helps plants to resist adverse environmental conditions. Our results showed that the size of stomatas were bigger in the *TRV:Gh4CL7* plants, suggesting a positive role of *Gh4CL7* in reducing transpiration rate that allows cotton to maintain a more favourable water balance, and effectively improves drought tolerance. This was supported by the observation of a higher water loss rate in the vitro leaves of the *TRV:Gh4CL7* plants than those of WT plants.

Lignin is the second largest polymer in plants after cellulose [59]. It provides mechanical support to plants by increasing cell wall hardness and enhancing compressive strength of cells [60–62]. We found that repressing the expression level of *Gh4CL7* in *G. hirsutum* reduced the lignin content and led to a reduction in drought resistance, consistent with the result of rice plants with a decreased lignin content being more prone to drought stress [63]. Studies in *Fraxinus mandshurica* also showed that decreased lignin content resulted in drought resistance reduction [24]. The hydrophobicity of lignin is thought to have an inhibitory effect on the transpiration of plant tissue under drought conditions [64], that could be

the reason for *Arabidopsis* plants overexpressing *Gh4CL7* that had an increased level of lignin content being more resistant to drought.

## Conclusions

The findings of this study demonstrate that the *Gh4CL7*-silencing cotton plants had an increased sensitivity of drought stress while overexpressing *Gh4CL7* enhanced tolerance of drought stress in *Arabidopsis*. *Gh4CL7* conferred tolerance to drought stress by increasing lignin content, improving the antioxidant system, closing stomata, and up-regulating the transcription levels of ABA-responsive genes. Although the exact mechanism of *Gh4CL7*-mediated drought tolerance is still yet to be uncovered, our results provide evidence for the role of *Gh4CL7* in combating drought stress.

## Methods

### Identification of the *4CL* family genes in *Gossypium hirsutum*

The annotated protein sequences of *G. hirsutum* [26] were downloaded from CottonGen (<https://www.cottongen.org/>). The hidden Markov model file corresponding to the AMP-binding domain (PF00501) was downloaded from the Pfam protein family database (<http://pfam.xfam.org/>) and used as query ( $P < 0.001$ ) [65] to search for the *4CL* genes in *G. hirsutum* with HMMER 3.0 [66]. The existence of the AMP-binding domain sequences was examined using the Pfam, SMART (<http://smart.embl-heidelberg.de/>), and National Center for Biotechnology Information (NCBI) Conserved Domains (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) databases [67, 68].

### Gene structure, conserved motif and promoter analyses

The length, molecular weight (MW), and isoelectric point ( $pI$ ) of the identified Gh4CL proteins were calculated using the ExPasy website tools (<http://web.expasy.org/protparam/>) [69]. Gene Structure Display Server 2.0 (GSDS, <http://gsds.cbi.pku.edu.cn/>) was used for intron and exon analysis [70]. The conserved motifs in the Gh4CL protein sequences were identified using the Multiple Expectation Maximization for Motif Elicitation (MEME) program (version 5.0.5, <http://meme-suite.org/tools/meme>) [71]. The potential *cis*-elements in the promoter sequences (up to 2000-bp upstream ATG) of *Gh4CL* genes were identified using the PlantCARE program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

### Phylogenetic tree, chromosomal distribution and syntenic relationship analyses

The multiple sequence alignment of Gh4CLs was done by Clustal X [72] and DNAMAN (version 5.2.2). The 4CL homologous protein sequences of *Arabidopsis thaliana* (At4CL1: OAP14948; At4CL2: OAP07084; At4CL3: AEE34324; At4CL4: AY376731; At4CL5: AY250839; At4CL7: AY376733; At4CL9: AF360250 At4CL11: AY376735), *Glycine max* (Gm4CL1: AF279267; Gm4CL2: AF002259; Gm4CL3: AF002258; Gm4CL4: X69955), *Rubus idaeus* (Ri4CL1: AF239687; Ri4CL2: AF239686; Ri4CL3: AF239685), *Populus tremuloides* (Pt4CL1: U12012; Pt4CL2: U12013), and *Isatis tinctoria* (li4CL1: ADG46006; li4CL2: KC430622; li4CL3: KC430623) were downloaded from the NCBI (<http://www.ncbi.nlm.nih.gov/>, accessed on 7 May 2018) and used for the phylogenetic tree analysis by using the neighbor joining method (NJ) in MEGA 6.0 [73] with 1000 repetitions for the bootstrap test.

All the *Gh4CL* genes were mapped to *G. hirsutum* chromosomes, based on their physical location information, using TBtools [74]. Two or more homologous/paralogous 4CL genes located at a chromosomal region of < 200 kb were considered to be generated by tandem duplication events [75]. Multiple Collinearity Scan toolkit (MCScanX) was used to analyze the gene duplication events with the default parameters [76]. Non-synonymous (Ka) and synonymous (Ks) substitutions in each paralogous *Gh4CL* gene pair were calculated using KaKs\_Calculator 2.0 [77].

## Vector construction and genetic transformation

To generate *Arabidopsis* overexpressing lines, the coding sequence of *Gh4CL7* was amplified using PrimeSTAR DNA polymerase (TaKaRa, Tokyo, Japan) with the gene-specific forward and reverse primers and ligated into the pCAMBIA2300 vector driven by the CaMV35S promoter. The expression vector pCAMBIA2300-Gh4CL7 was transformed into *Agrobacterium tumefaciens* strain GV3101. *Arabidopsis* ecotype Col-0 (wild-type, provided by Microbiology Institute of Chinese Academy of Sciences) was used in genetic transformation by the floral dip method [78]. The harvested T0 generation seeds were selected on 1/2 Murashige and Skoog (MS) medium with 50 mg/L kanamycin, and the resistant plants were further validated by PCR. Single-copy lines with a segregation ratio of 3:1 were selected and planted until T3 generation. The transgenic *Arabidopsis* overexpressing lines were grown in a growth chamber with 16 h light/8 h dark scheme, and the growth temperature was set at around 23 °C [79].

The pTRV1 and pTRV2 VIGS vectors were kindly provided by Prof. Longfu Zhu of [Huazhong Agricultural University](#). A 287-bp fragment from the coding sequence of *Gh4CL7* was amplified from cDNA of *G. hirsutum* cultivar stem and subsequently cloned into pTRV2 using *Bam*H1 and *Kpn*1 double digestions to generate the *TRV:Gh4CL7* vector. After PCR and double digestion confirmation, the *TRV:Gh4CL7* construct was transformed into *A. tumefaciens* strain GV3101 by electroporation. *G. hirsutum* cv. Junmian-1 (provided by the Cotton Research Institute of Shihezi University) was used in VIGS and was grown in pots filled with soil mix (3:1, humus:vermiculite) and placed in growth room at 24 ± 1 °C with 16 h light/8 h dark (200 µmol/m<sup>2</sup>/s photon flux density). VIGS was done as previously described by Xiong and colleagues [79].

# Drought tolerance assays

Polyethylene glycol (PEG) and natural drought treatments were conducted to investigate the function of *Gh4CL7* in response to osmotic and drought stresses. For the PEG treatments, *TRV:Gh4CL7* (VIGS) and *TRV:00* (control) plants at three-leaf stage were subjected to stress by 20% PEG6000 (w/v). Leave samples were collected at 0 (ck), 1, 3, 6, 9, 12, 24 h after treatment for RNA extraction. For the natural drought treatments, *TRV:Gh4CL7* (VIGS) and *TRV:00* (control) plants were not irrigated for three weeks, followed by re-watering once. After 15 days of water deprivation, plant leaves were collected for RNA extraction and determination of physiological parameters.

For analyses of seed germination and root elongation, seeds or 3-day-old seedlings of *Gh4CL7* - overexpressing *Arabidopsis* lines and WT were grown on 1/2 MS plates supplemented with 0 (control), 200, 300 mM mannitol for six days. In the vegetative growth stage, 7 days-old transgenic *Arabidopsis* and WT plants were transplanted into soil and watered once every 3 days for 2 weeks, then kept without irrigation for 2 weeks, followed by re-watering once. After 10 days of drought treatment, the rosette leaves were collected for RNA extraction and determination of physiological parameters.

## Determination of drought stress-related physiological parameters

The thiobarbituric acid (TBA) colorimetric method was used to measure the content of malondialdehyde (MDA) according to the instruction of the malondialdehyde quantification kit (Suzhou Comin Biochemistry Co. Ltd, Su Zhou, China).  $H_2O_2$  concentrations were determined by using the  $H_2O_2$  determination kit (Suzhou Comin Biochemistry Co. Ltd, Su Zhou, China) by following the manufacturer's instructions. Measurement of the activities of antioxidative enzymes was performed using 0.1 g leave sample according to the instructions of the POD, SOD and CAT Assay Kit (Suzhou Comin Biochemistry Co. Ltd, Su Zhou, China). Total contents of chlorophyll were calculated according to the method described by Porra and colleagues [80]. Absorbances at different wavelengths were measured by the U-5100 UV/VIS spectrophotometer (HITACHI, Tokyo, Japan).

## Measurements of weight lose relative (WLR), relative water content (RWC) and stomatal aperture

For water loss assays, leaves from *G. hirsutum* and *Arabidopsis* were immediately weighed and placed in a growth room at room temperature with a humidity level of about 60%. The leaves were weighed once per hour. Water loss relative (WLR) was estimated as the percentage of fresh weight lost relative to the initial fresh weight [81]. For measurement of the relative water content (RWC), fresh leaves were detached from plants and their fresh weights (FW) were immediately recorded. Then, the leaves were placed in distilled water for 8 hours at 25°C in the dark and measured the turgid weight (TW). Dry weights (DW)

were recorded by drying samples at 65 °C until constant weight. RWC was calculated as  $(FW - DW)/(TW - DW) \times 100$  [82].

Stomatal pore area and size were determined using the rapid imprinting technique [83]. The abaxial leaf surfaces were covered with transparent nail polish and air dried at room temperature. The nail polish imprints were made into temporary slices and photographed by Zeiss microscope (SteREO Discovery.V20, Germany) with 300× magnifications. The length and width of stomatal pores were measured using the Image J software and the relative aperture area was calculated based on the ratio of width to length.

## Lignin content measurement and histochemical staining

Lignin content was determined by the acetyl bromide method [84] and the phloroglucinol-HCl color-developing method. Transections of stem (above the cotyledons) of the *Gh4CL7*-silencing cotton plants at 28 days after sowing were made by hand cutting using razor blades. The transverse sections were immersed in 1 mL of 1 M hydrochloric acid for 3 min, followed by transferring into 1 ml 10% phloroglucinol ethanol-hydrochloric solution for 1 min, and visualized immediately under the microscope (Zeiss, SteREO Discovery.V20, Germany) for photographing [85].

## RNA extraction and quantitative real-time PCR analysis

To investigate gene expression patterns, total RNA was extracted from leaves of *G. hirsutum* and *Arabidopsis* with the EASYspin Plus plant RNA kit (Aidlab, Beijing, China). RNA was reverse transcribed into cDNA using the M-mlv reverse transcript system (TAKARA, Da Lian, China). The quantitative real-time polymerase chain reaction (qRT-PCR) was performed using the Power SYBR Green PCR Master Mixture (Roche, Rotkreuz, Switzerland) on a Light Cycler® 480 II system (Roche, Rotkreuz, Switzerland) under the following conditions: initial pre-incubation at 95°C for 5 min, followed by 40 cycles at 94°C for 10 s, 59°C for 10 s, and 72°C for 10 s. The relative expression level of genes was analyzed by the  $2^{-\Delta\Delta Ct}$  method. The results were presented as the mean of three biological replications. The *G. hirsutum histone3* gene and *Arabidopsis EF-1a* gene were used as the reference genes. All the primers used in this study were designed using the NCBI primer designing tool (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>, accessed on 27 August 2018) and listed in Additional file 2: Table S3.

## Statistical analyses

Statistical analyses and data plotting were performed using SPSS and Graphpad Prism 5, respectively. \*\* and \* represent significant differences at  $P < 0.01$  and  $P < 0.05$ , respectively.

## Abbreviations

4CL: 4-coumarate-CoA ligases; VIGS: Virus-induced gene silencing; OE: overexpression; RWC: Relative water content; MDA: Malondialdehyde; H<sub>2</sub>O<sub>2</sub>: Hydrogen peroxide; ABA: abscisic acid; PEG: polyethylene glycol; DPA: Day post anthesis; qRT-PCR: quantitative reverse transcription polymerase chain reaction; CAT: Catalase; POD: Peroxidase; SOD: Superoxide dismutase; ROS: Reactive oxygen species; WT: Wild type; TRV:Gh4CL7: *Gh4CL7* VIGS plants; TRV:00: empty vector VIGS plants

## Declarations

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Availability of data and materials

All Gene ID, annotation files and RNA-Seq data could be obtained from CottonGen (<https://www.cottongen.org>). All other data generated or analyzed during this study are included in this manuscript.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the National Natural Science Foundation of China (Grant No: 31906438, 31360347), the National Key Research and Development Program of China (Grant No: 2016YFD0100200) and the Breeding Program of Shihezi University (Grant No:YZZX201601). The funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

## Authors' contributions

SS and XX designed the experiments. SS, XX and XZ performed the experiments, analysed data and prepared the manuscript. HF contributed with valuable discussions. QZ, YL and JS read and revised the manuscript. All authors provided helpful discussions and approved its final version.

# Acknowledgements

Not applicable.

## Reference

1. Hussain M, Farooq S, Hasan W, Ul-Allah S, Tanveer M, Farooq M, Nawaz A. Drought stress in sunflower: physiological effects and its management through breeding and agronomic alternatives. *Agricultural Water Management*. 2018;201:152–66.
2. Araújo AE, Silva CAD, Azevedo DMP et al. Cultivo do algodão irrigado. Série Documentos, Sistemas de Produção, 3rd, Versão eletrônica, 2003. p. 1678–8710.
3. Pasapula V, Shen GX, Kuppu S, Paez-Valencia J, Mendoza M, Hou P, Chen J, Qiu XY, Zhu LF, Zhang XL, Auld D, Blumwald E, Zhang H, Gaxiola R, Payton P. Expression of an Arabidopsis vacuolar H<sup>+</sup>-pyrophosphatase gene (AVP1) in cotton improves drought- and salt tolerance and increases fibre yield in the field conditions. *Plant Biotechnology Journal*. 2011;9:88–99.
4. Hu Y, Li WC, Xu YQ, Li GJ, Liao Y, Fu FL. Differential expression of candidate genes for lignin biosynthesis under drought stress in maize leaves. *J Appl Genet*. 2009;50:213–23.
5. Geng D, Chen P, Shen X, Zhang Y, Li X, Jiang L, Niu C, Zhang J, Huang X, Ma F, Guan Q. MdMYB88 and MdMYB124 enhance drought tolerance by modulating root vessels and cell walls in apple. *Plant Physiol*. 2018;178:1296–309.
6. Hu Q, Min L, Yang XY, Jin SX, Zhang L, Li YY, Ma YZ, Qi XW, Li DQ, Liu HB, Lindsey K, Zhu LF, Zhang XL. Laccase GhLac1 modulates broad-spectrum biotic stress tolerance via manipulating phenylpropanoid pathway and jasmonic acid synthesis. *Plant Physiol*. 2018;176:1808–23.
7. Zhang K, Cui H, Cao S, Yan L, Li M, Sun Y. Overexpression of CrCOMT from *Carex rigescens* increases salt stress and modulates melatonin synthesis in *Arabidopsis thaliana*. *Plant Cell Reports*. 2019; <https://doi.org/10.1007/s00299-019-02461-7>.
8. Ehltting JR, Büttner DB, Wang Q, Douglas CJ, Somssich IE, Kombrink E. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana*, represent two evolutionarily divergent classes in angiosperms. *Plant J*. 1999;19:9–20.
9. Vogt T. Phenylpropanoid biosynthesis. *Mol Plant*. 2010;3:2–20.
10. Lavhale SG, Kalunke RM, Giri S, Structural AP. functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta*. 2018;248:1063–78.
11. Costa MA, Bedger DL, Moinuddin SGA et al. Characterization in vitro and in vivo of the putative multigene 4-coumarate: CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl

alcohol derivative formation. *Phytochemistry*. 2005;66:2072–91.

12. Gui JS, Shen JH, Li LG. Functional Characterization of Evolutionarily Divergent 4-Coumarate: Coenzyme A Ligases in Rice. *Plant Physiology*. 2011;157:574–86.

13. Sun HY, Li Y, Feng SQ, Zou WH, Guo K, Fan CF, Si SL, Peng LC. Analysis of five rice 4-coumarate: coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochem Biophys Res Commun*. 2013;430:1151–6.

14. Uhlmann A, Ebel J. Molecular cloning and expression of 4-coumarate:Coenzyme a ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack. *Plant Physiol*. 1993;102:1147–56.

15. Shi R, Sun YH, Li Q, Heber S, Sederoff R, Chiang VL. Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol*. 2010;51:144–63.

16. Souza CA, Barbazuk B, Ralph SG, Bohlmann J, Hamberger B, Douglas CJ. Genome-wide analysis of a land plant-specific acyl:coenzyme A synthetase (ACS) gene family in *Arabidopsis*, poplar, rice and *Physcomitrella*. *New Phytol*. 2008;179:987–1003.

17. Hu WJ, Kawaoka A, Tsai CJ, Lung J, Osakabe K, Ebinuma H, Chiang VL. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:coa ligase genes in aspen (*populus tremuloides*). *P Natl Acad Sci USA*. 1998;95:5407–12.

18. Raes J, Rohde A, Christensen JH, VandePeer Y, Boerjan W. Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant Physiol*. 2003;133:1051–71.

19. Zhang CH, Ma T, Luo WC, Xu JM, Liu JQ, Wan DS. Identification of 4CL Genes in Desert Poplars and Their Changes in Expression in Response to Salt Stress. *Genes*. 2015;6:901–17.

20. Naik P, Wang JP, Sederof R et al. Assessing the impact of the 4CL enzyme complex on the robustness of monolignol biosynthesis using metabolic pathway analysis. *PLoS One*. 2018;13:e0193896.

21. Li Y, Kim JI, Pysh L, Chapple C. Four isoforms of *arabidopsis thaliana* 4-coumarate: coa ligase (4cl) have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol*. 2014;169:2409–15.

22. Gao S, Yu HN, Xu RX, Cheng AX, Lou HX. Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagi ochasma appendiculatum*. *Phytochemistry*. 2015;111:48–58.

23. Di P, Hu YS, Xuan HJ, Xiao Y, Chen JF, Zhang L, Chen WS. Characterisation and the expression profile of 4-coumarate:coa ligase (li4CL) from hairy roots of *isatis indigotica*. *Afr J Pharm Pharmacol*. 2012;6:2166–75.



- 24.Chen XH, Wang HT, Li XY, Ma K, Zhan YG, Zeng FS. Molecular cloning and functional analysis of 4-Coumarate:CoA ligase 4(4CL-like 1) from *Fraxinus mandshurica* and its role in abiotic stress tolerance and cell wall synthesis. *BMC Plant Biology*. 2019;19:231–47.
- 25.Cannon SB, Mitr A, Baumgarten A, Young ND, May G. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*. *BMC Plant Biol*. 2004;4:10–20.
- 26.Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Yves VDP, Pierree R, Stephane R. Plantcare, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res*. 2002;30:325–7.
- 27.Zhang T, Hu Y, Jiang W, Fang L, Chen ZJ. Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat Biotechnol*. 2015;33:531–7.
- 28.Rao GD, Pan X, Xu F, Zhang YZ, Cao S, Jiang XG, Lu H. Divergent and overlapping function of five 4-Coumarate: coenzyme A ligases from *Populus tomentosa*. *Plant Mol Biol Rep*. 2015;33:841–54.
- 29.Lindermayr C, Möllers B, Fliegmann J, Uhlmann A, Lottspeich F, Meimberg H, Ebel J. Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. *Eur J Biochem*. 2002;269:1304–15.
- 30.Hamberger B, Hahlbrock K. The 4-coumarate:coa ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *P Natl Acad Sci USA*. 2004;101:2209–14.
- 31.Chen HC, Song J, Williams CM, Shuford CM, Liu J, Wang JP, Li Q, Shi R, Gokce E, Ducoste J, Muddiman DC, Sederoff RR, Chiang VL. Monolignol pathway 4-coumaric acid:coenzyme A ligases in *Populus trichocarpa*: novel specificity, metabolic regulation, and simulation of coenzyme A ligation fluxes. *Plant Physiol*. 2013;161:1501–16.
- 32.Choudhary EK, Choi B, Cho BK, Kim JB, Park SU, Natarajan S, Lim HS, Bael H. Regulation of 4CL, encoding 4-coumarate: coenzyme A ligase, expression in kenaf under diverse stress conditions. *Plant Omics J*. 2013;6:254–62.
- 33.Sutela S, Hahl T, Tiimonen H, Aronen T, Ylloja T, Laakso T, Saranpää P, Chiang V, Julkunen-Tiitto R, Häggman H. Phenolic compounds and expression of 4CL genes in silver birch clones and Pt4CL1a lines. *Plos One*. 2014;9:e114434.
- 34.Lee D, Douglas CJ. Two divergent members of a tobacco 4-Coumarate: coenzyme A ligase (4CL) gene family (cDNA structure, gene inheritance and expression, and properties of recombinant proteins). *Plant Physiol*. 1996;112:193–205.
35. Shinde BA, Dholakia BB, Hussain K, Panda S, Meir S, Rogachev I, Aharoni A, Giri AP, Kamble AC. Dynamic metabolic reprogramming of steroidal glycol-alkaloid and phenylpropanoid biosynthesis may

- impart early blight resistance in wild tomato (*Solanum arcanum* Peralta). *Plant Mol Biol.* 2017;95:411–23.
- 36.Ehlting J, Shin JJ, Douglas CJ. Identification of 4-coumarate:coenzyme A ligase (4CL) substrate recognition domains. *Plant J.* 2001;27:455–65.
- 37.Schneider K, Hövel K, Witzel K, Hamberger B, Schomburg D, Kombrink E, Stuible H-P. The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. *Proc Natl Acad Sci USA.* 2003;100:8601–6.
- 38.Rose AB. Intron-mediated regulation of gene expression. *Curr Top Microbiol Immunol.* 2008;326:277–90.
- 39.Xu G, Guo C, Shan H, Kong H. Divergence of duplicate genes in exon-intron structure. *Proc Nat Acad Sci USA.* 2012;109:1187–92.
- 40.Zhao P, Wang DD, Wang RQ, Kong NN, Zhang C, Yang CH, Wu WT, Ma HL, Chen Q. Genome-wide analysis of the potato Hsp20 gene family: identification, genomic organization and expression profiles in response to heat stress. *BMC Genomics.* 2018;19:61–73.
- 41.Xu J, Xu XY, Tian LL, Wang GL, Zhang XY, Guo WZ. Discovery and identification of candidate genes from the chitinase gene family for *Verticillium dahliae* resistance in cotton. *Scientific Reports.* 2016; 6:29022.
- 42.Liu Z, Ge XY, Yang ZR, Zhang CJ, Zhao G, Chen EY, Liu J, Zhang XY, Li FG. Genome-wide identification and characterization of SnRK2gene family in cotton (*Gossypium hirsutum*L.). *BMC Genetics.* 2017;18:54–68.
- 43.Turner JG, Ellis C, Devoto A. The jasmonate signal pathway. *Plant Cell.* 2002;14:S153–S164.
- 44.Browse J and Howe GA. New weapons and a rapid response against insect attack. *Plant Physiol.* 2008;146:832–8.
- 45.Xu WR, Yu YH, Ding JH, Hua ZY, Wang YJ. Characterization of a novel stilbene synthase promoter involved in pathogen and stress inducible expression from Chinese wild *Vitis pseudoreticulata*. *Planta.* 2010;231:475–87.
- 46.Li W, Cui X, Meng ZL, Huang XH, Xie Q, Wu H, Jin HL, Zhang DB, Liang WQ. Transcriptional regulation of Arabidopsis MIR168a and ARGONAUTE1 homeostasis in abscisic acid and abiotic stress responses. *Plant Physiol.* 2012;158:1279–92.
- 47.Hossain MA, Bhattacharjee S, Armin S-M, Qian PP, Xin W, Li HY, Burritt DJ, Fujita M, Tran LSP. Hydrogen peroxide priming modulates abiotic oxidative stress tolerance: insights from ROS detoxification and scavenging. *Front Plant Sci.* 2015;6:420–39.

- 48.Jain G, Gould KS. Are betalain pigments the functional homologues of anthocyanins in plants? *Environ Exp Bot.* 2015;119:48–53.
- 49.Parkhi V, Kumar V, Sunilkumar G, Campbell LM, Singh NK, Rathore KS. Expression of apoplastically secreted tobacco osmotin in cotton confers drought tolerance. *Mol Breed.* 2009;23:625–39.
- 50.Sharma P, Jha AB, Dubey RS, Pessarakli M. Reactive oxygen species, oxidative damage, and antioxidative defense mechanism in plants under stressful conditions. *J Exp Bot.* 2012;12:1–26.
- 51.Ullah A, Sun H, Hakim YX, Zhang X. A novel cotton WRKY gene, GhWRKY6-like, improves salt tolerance by activating the ABA signaling pathway and scavenging of reactive oxygen species. *Physiol Plant.* 2018;162:439–54.
- 52.Comas LH, Becker SR, Cruz VMV, Byrne PF, Dierig DA. Root traits contributing to plant productivity under drought. *Front Plant Sci.* 2013;4:442–58.
- 53.Xiong L, Wang RG, Mao G, Koczan JM. Identification of drought tolerance determinants by genetic analysis of root response to drought stress and abscisic acid. *Plant Physiol.* 2006;142:1065–74.
- 54.Montillet JL, Leonhardt N, Mondy S, Tranchimand S, Rumeau D, Boudsocq M, Garcia AV, Douki T, Bigeard J, Laurie re C, Chevalier A, Castresana C, Hirt H. An abscisic acid-independent oxylipin pathway controls stomata closure and immune defense in Arabidopsis. *PLoS Biol.* 2013;11:e1001513.
- 55.Rowe JH, Topping JF, Liu J, Lindsey K. Abscisic acid regulates root growth under osmotic stress conditions via an interacting hormonal network with cytokinin, ethylene and auxin. *New Phytol.* 2016;211:225–39.
- 56.Skriver K, Mundy J. Gene expression in response to abscisic acid and osmotic stress. *Plant Cell.* 1990;2:503–12.
- 57.Chandler PM, Robertson M. Gene expression regulated by abscisic acid and its relation to stress tolerance. *Annu. Rev. Plant Physiol. Plant Mol Biol.* 1994;45:113–41.
- 58.Leung J, Giraudat J. Abscisic acid signal transduction. *Annu Rev Plant Physiol Plant Mol Biol.* 1998;49:199–222.
- 59.Penning BW, Hunter CT, Tayengwa R, Eveland AL, Dugard CK, Olek AT, Vermerris W, Koch KE, McCarty DR, Davis MF, Thomas SR, McCann MC, Carpita NC. Genetic resources for maize cell wall biology. *Plant Physiol.* 2009;151:1703–28.
- 60.Campbell M, Sederoff R. Variation in lignin content and composition (mechanisms of control and implications for the genetic improvement of plants). *Plant Physiol.* 1996;110:3–13.

- 61.Vance C, And TK, Sherwood R. Lignification as a mechanism of disease resistance. *Annu Rev Phytopathol.* 2003;18:259–88.
- 62.Lin CY, Wang JP, Li QZ, Chen HC, Liu J, Loziuk P, Song J. 4-Coumaroyl and caffeoyl shikimic acids inhibit 4-coumaric acid: coenzyme a ligases and modulate metabolic flux for 3-hydroxylation in monolignol biosynthesis of *Populus trichocarpa*. *Mol Plant.* 2015;8:176–87.
- 63.Li WQ, Zhang MJ, Gan PF, Qiao L, Yang SQ, Miao H, Wang GF, Zhang MM, Liu WT, Li HF, Shi CH, Chen KM. CLD1/SRL1 modulates leaf rolling by affecting cell wall formation, epidermis integrity and water homeostasis in rice. *Plant J.* 2017;92:904–23.
- 64.Jordan R. Structural changes and associated reduction of hydraulic conductance in roots of *Sorghum bicolor* L. following exposure to water deficit. *Plant Physiol.* 1992;99:203–12.
- 65.Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR. The pfam protein families database. *Nucleic Acids Res.* 2014;42:222–30.
- 66.Finn RD, Clements J, Eddy SR. Hmmer web server: interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39:29–37.
- 67.Aron MB, Myra KD, Noreen RG, Shennan L, Farideh C, Lewis YG. Cdd: ncbi's conserved domain database. *Nucleic Acids Res.* 2015;43:D222.
- 68.Letunic I, Bork P. 20 years of the smart protein domain annotation resource. *Nucleic Acids Res.* 2018;46:D493–D496.
- 69.Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker J.M, editor. *The Proteomics Protocols Handbook*. Totowa:New Jersey; 2005. p. 571–607.
- 70.Guo AY, Zhu QH, Chen X, Luo JC. GSDS: a gene structure display server. *Hereditas.* 2007;29:1023–26.
- 71.Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren JY, Li WW, Noble WS. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–W208.
72. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997;25:4876–82.
- 73.Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. Mega6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30:2725–9.
- 74.Chen CJ, Xia R, Chen H, He YH. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. *BioRxiv.* 2018;<http://doi.org/10.1101/289660>.

- 75.Holub EB. The arms race is ancient history in arabidopsis, the wildflower. *Nat Rev Genet.* 2001;2:516–27.
- 76.Wang YP, Tang HB, Debarry JD, Tan X, Li JP, Wang XY, Lee TH, Jin HZ, Marler B, Guo H, Kissinger JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research.* 2012;40:e49.
- 77.Wang D, Zhang Y, Zhang Y, Zhu J, Yu J. KaKs\_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinf.* 2010;8:77–80.
- 78.Clough SJ, Bent AF. Floral dip: a simplified method for *Agrobacterium*-mediated transformat of *Arabidopsis thaliana*. *Plant J.* 1998;16:735–43.
- 79.Xiong XP, Sun SC, Li YJ, Zhang XY, Sun J, Xue F. The cotton WRKY transcription factor GhWRKY70 negatively regulates the defense response against *Verticillium dahliae*. *The Crop Journal.* 2019;7:393–402.
- 80.Porra RJ, Thompson WA, Kriedemann PE. Determination of accurate extinction coefficients and simultaneous equations for assaying chlorophylls a and b extracted with four different solvents:verification of the concentration of chlorophyll standards by atomic absorption spectroscopy. *Biochimica et Biophysica Acta.* 1989;975:384–94.
- 81.Lefebvre V, North H, Frey A, Sotta B, Seo M, Okamoto M, Nambara E, Marion-Poll A. Functional analysis of *Arabidopsis* NCED6 and NCED9 genes indicates that ABA synthesized in the endosperm is involved in the induction of seed dormancy. *Plant J.* 2006;45:309–19.
- 82.Sade N, Vinocur BJ, Diber A, Shatil A, Ronen G, Nissan H, Wallach R, Karchi H, Moshelion M. Improving plant stress tolerance and yield production: is the tonoplast aquaporin SITIP2; 2 a key to isohydric to anisohydric conversion?. *New Phytol.* 2009;181:651–61.
- 83.Geisler M, Nadeau J, Sack FD. Oriented asymmetric divisions that generate the stomatal spacing pattern in arabidopsis are disrupted by the too many mouths mutation. *Plant Cell.* 2000;12:2075–86.
- 84.Morrison I. A semi-micro method for the determination of lignin and its use in predicting the digestibility of forage crops. *J Sci Food Agric.* 1972;23:455–63.
- 85.Pomar F, Merino F, Barceló AR. O–4-Linked coniferyl and sinapyl aldehydes in lignifying cell walls are the main targets of the Wiesner (phloroglucinol-HCl) reaction. *Protoplasma.* 2002;220:17–28.

## Additional File

*Additional file 1:*

*Fig. S1* Alignment of multiple Gh4CL and selected At4CL domain amino acid sequences. Multiple sequence alignment was performed using Clustal X. Box I and Box II represent the two conserved domains of the Gh4CL proteins.

*Fig. S2* Relative expression levels of *Gh4CL7* in plants infiltrated with *TRV:00* and *TRV:Gh4CL7* (n = 5). Total RNA was extracted from leaves at 2 weeks post-infiltration. Transcript levels were determined by qRT-PCR using *GhUBQ7* as control.

*Fig. S3* Silencing of the endogenous magnesium chelatase subunit I gene (*GhCHLI*) in cotton through VIGS. The leaf bleaching phenotype was observed two weeks after infiltration in TRV: GhCHLI plants. *Fig. S4* Transcription levels of stress responsive genes in CK and Gh4CL7-silencing cotton plants. The normal condition plants were used as controls. *GhUBQ7* gene was used as an internal control. All the gene expressions were normalized to the corresponding transcript levels in CK plants at normal condition. All the data represent mean±SE for three biological replications.

*Fig. S5* Transcription levels of stress responsive genes in transgenic and WT *Arabidopsis* plants. The normal condition plants were used as controls. *AtEF-La* gene was used as an internal control. All the gene expressions were normalized to the corresponding transcript levels in WT plants at normal condition. All the data represent mean±SE for three biological replications.

*Additional file 2:*

*Table S1* Ka, Ks, Ka/Ks values for the Gh4CL paralogous gene pairs.

*Table S2* The 20 motifs information of Gh4CL proteins.

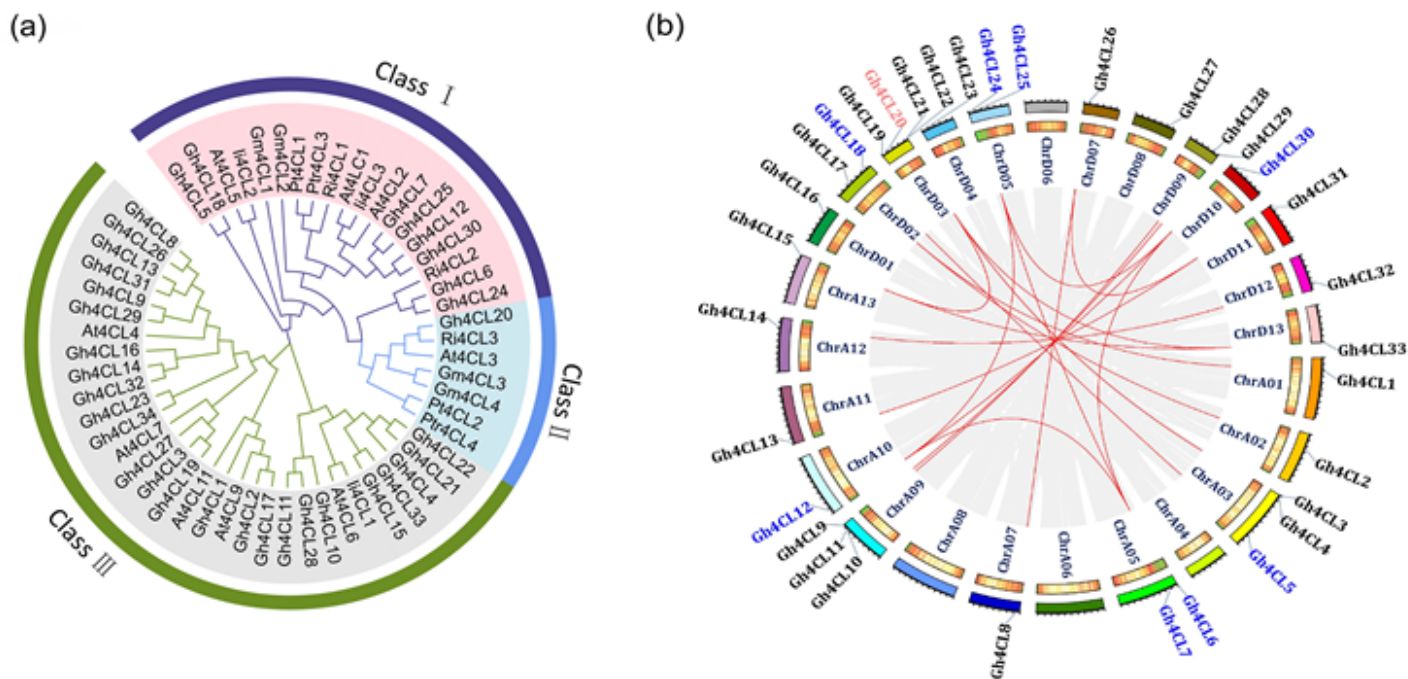
*Table S3* The primers used in qRT-PCR analysis.

# Table

**Table 1:** Characteristics of the 34 *Gh4CL* genes in *G. hirsutum*.

Genes	Gene ID number	Length/aa	MW/KDa	pI	subcellular localization
<i>Gh4CL1</i>	Gh_A01G0870	552	60.36	8.66	mitochondrial inner membrane
<i>Gh4CL2</i>	Gh_A02G0862	548	59.56	8.70	microbody (peroxisome)
<i>Gh4CL3</i>	Gh_A03G1962	458	50.34	6.35	chloroplast thylakoid membrane
<i>Gh4CL4</i>	Gh_A03G0249	517	57.08	8.32	plasma membrane
<i>Gh4CL5</i>	Gh_A03G1091	555	61.09	5.73	plasma membrane
<i>Gh4CL6</i>	Gh_A05G3997	543	59.56	5.67	plasma membrane
<i>Gh4CL7</i>	Gh_A05G1188	543	59.51	5.84	endoplasmic reticulum (membrane)
<i>Gh4CL8</i>	Gh_A07G0468	545	60.04	6.00	plasma membrane
<i>Gh4CL9</i>	Gh_A09G2180	557	60.84	6.80	Microbody (peroxisome)
<i>Gh4CL10</i>	Gh_A09G1370	276	30	5.95	endoplasmic reticulum (membrane)
<i>Gh4CL11</i>	Gh_A09G1371	572	63.38	9.76	plasma membrane
<i>Gh4CL12</i>	Gh_A10G0456	543	59.61	5.30	endoplasmic reticulum (membrane)
<i>Gh4CL13</i>	Gh_A11G0333	550	60.16	7.65	microbody (peroxisome)
<i>Gh4CL14</i>	Gh_A12G1362	521	56.42	8.30	microbody (peroxisome)
<i>Gh4CL15</i>	Gh_A13G2028	545	60.08	8.97	plasma membrane
<i>Gh4CL16</i>	Gh_D01G1584	501	54.36	7.16	plasma membrane
<i>Gh4CL17</i>	Gh_D02G0989	576	62.98	8.66	microbody (peroxisome)
<i>Gh4CL18</i>	Gh_D02G1514	555	61.10	5.80	plasma membrane
<i>Gh4CL19</i>	Gh_D03G1840	572	63.08	6.63	chloroplast thylakoid membrane
<i>Gh4CL20</i>	Gh_D03G0479	573	62.04	5.54	plasma membrane
<i>Gh4CL21</i>	Gh_D03G1317	452	49.64	8.40	plasma membrane
<i>Gh4CL22</i>	Gh_D03G1318	512	56.36	8.79	plasma membrane
<i>Gh4CL23</i>	Gh_D04G0054	568	61.94	6.17	microbody (peroxisome)
<i>Gh4CL24</i>	Gh_D05G3934	540	59.26	5.90	endoplasmic reticulum (membrane)
<i>Gh4CL25</i>	Gh_D05G1366	543	59.42	5.71	endoplasmic reticulum (membrane)
<i>Gh4CL26</i>	Gh_D07G0533	529	58.26	6.30	plasma membrane
<i>Gh4CL27</i>	Gh_D08G1670	528	57.83	5.96	microbody (peroxisome)
<i>Gh4CL28</i>	Gh_D09G1372	615	67.73	6.44	cytoplasm
<i>Gh4CL29</i>	Gh_D09G2385	557	60.94	6.50	microbody (peroxisome)
<i>Gh4CL30</i>	Gh_D10G0473	543	59.52	5.42	endoplasmic reticulum (membrane)
<i>Gh4CL31</i>	Gh_D11G0389	550	60.01	7.19	microbody (peroxisome)
<i>Gh4CL32</i>	Gh_D12G1488	547	59.54	8.74	microbody (peroxisome)
<i>Gh4CL33</i>	Gh_D13G2431	545	60.04	8.97	plasma membrane
<i>Gh4CL34</i>	Gh_Sca008083G01	568	61.77	6.09	microbody (peroxisome)

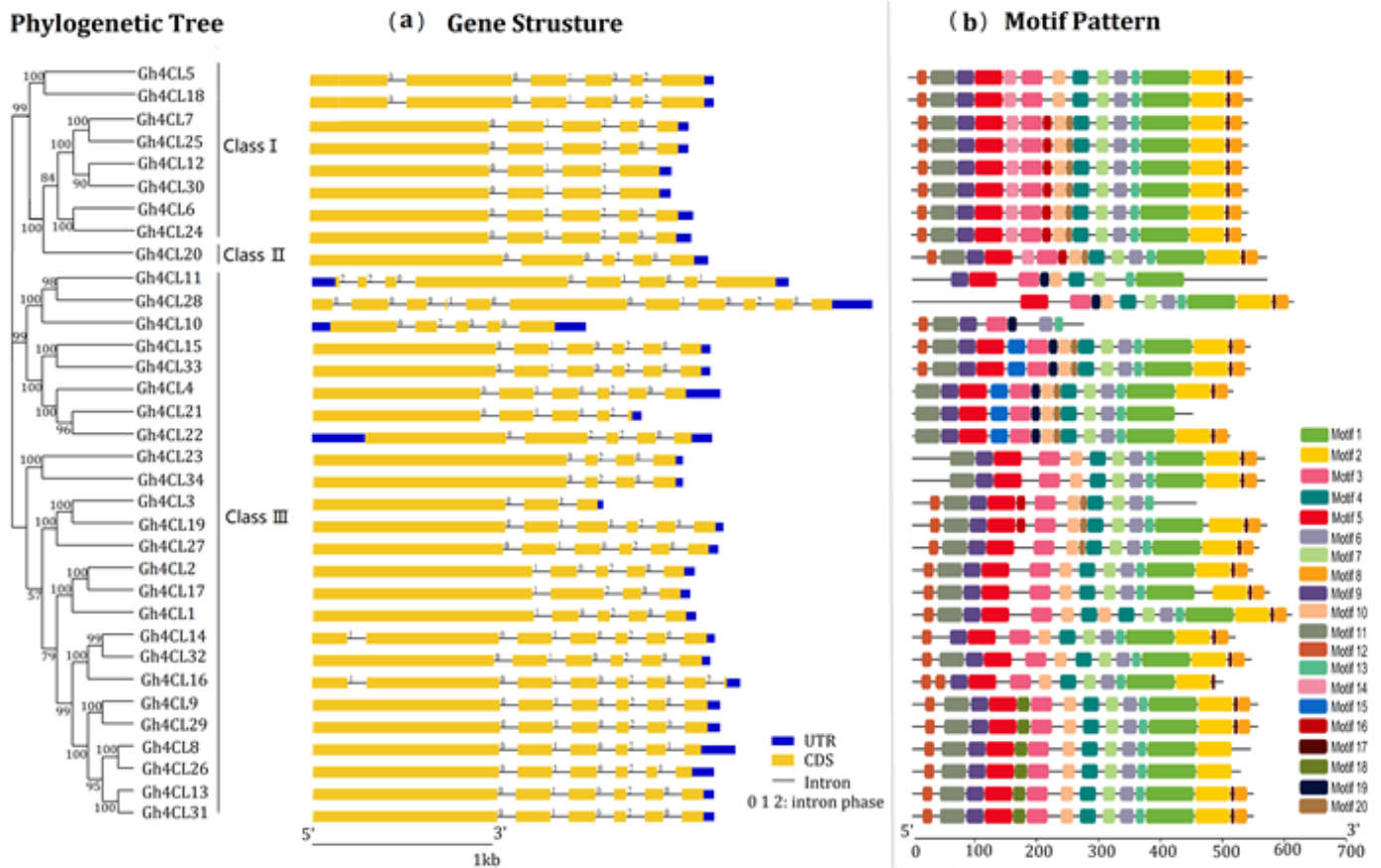
## Figures



**Figure 1**

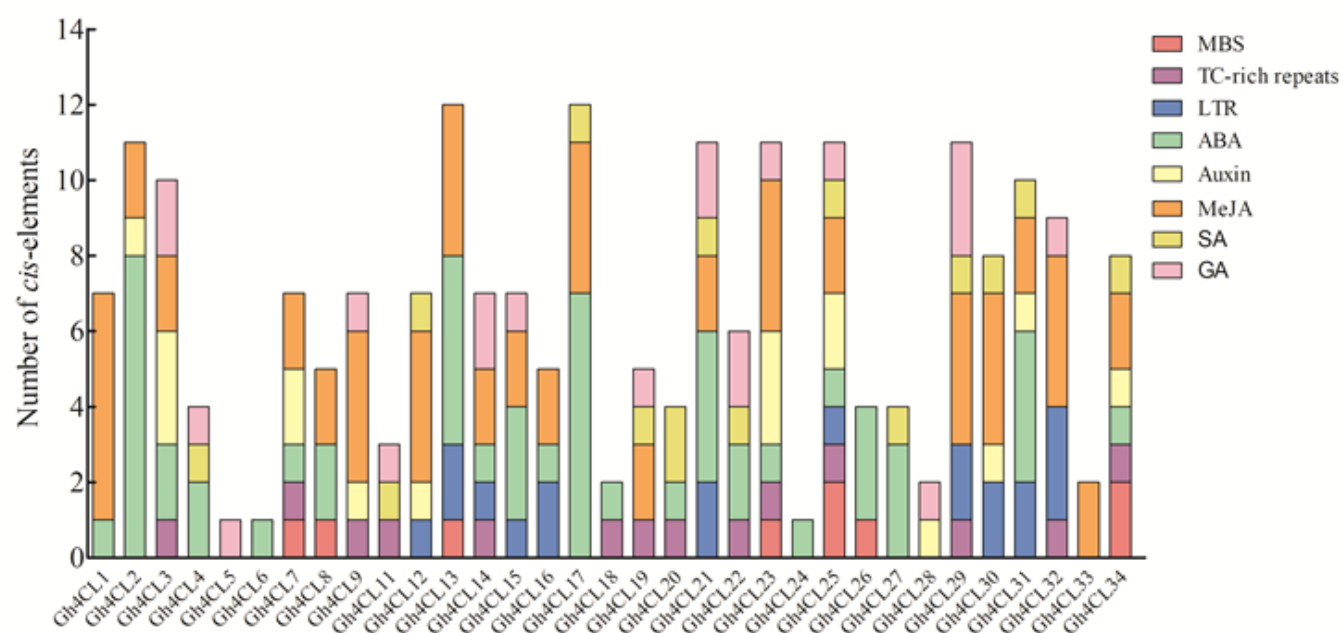
Phylogenetic, chromosomal distribution and interchromosomal relationships of the Gh4CL genes. a The neighbor-joining phylogenetic tree was generated using 4CL protein sequences from *G. hirsutum* and other six plants. b Chromosomal distribution of Gh4CL genes. The gray lines indicate all syntenic blocks in the genome of *G. hirsutum* and the red lines indicate interchromosomal relationships of Gh4CLs. The first circle represents individual chromosomes length. The second circle represents the gene density of each chromosome by color-coded short bars. The chromosome number were shown inside the circle.





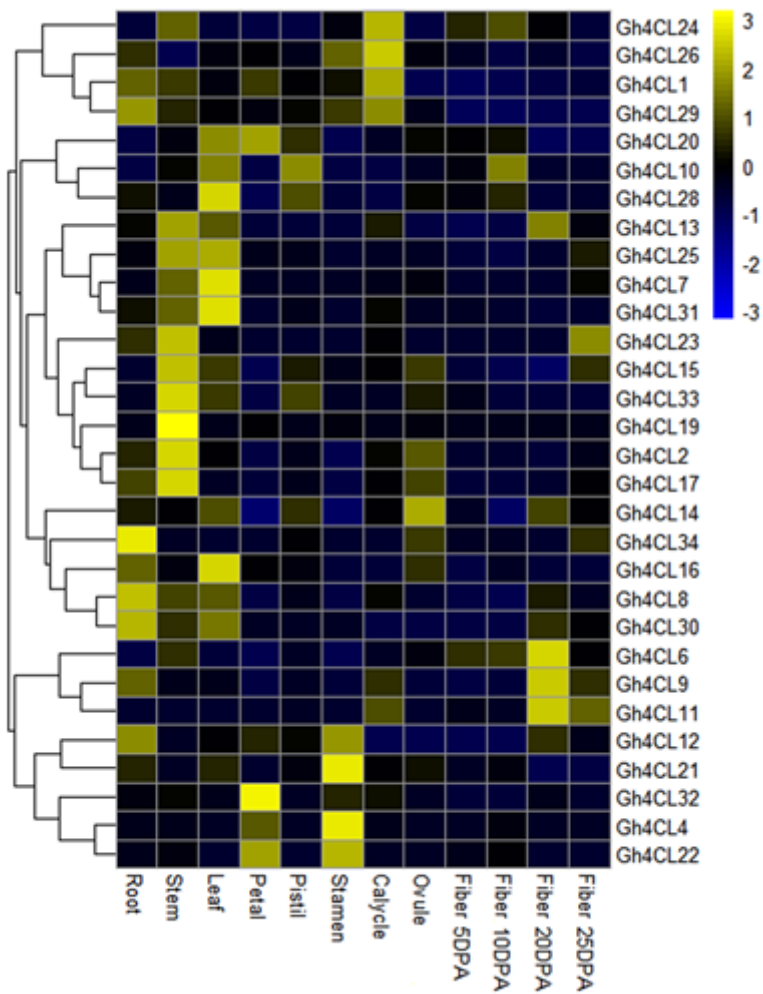
**Figure 2**

Gene structure and motif pattern of the Gh4CL genes. a Structural analysis. The left panel shows the neighbor-joining phylogenetic tree based on the amino acid sequences of the Gh4CLs. The classes I, II and III were marked correspondingly. The right panel shows the exon-intron structure of each Gh4CL gene with exons showing in orange boxes, introns in black lines between exons, and upstream/downstream (UTRs) in blue boxes. The number indicates the phase of the corresponding introns. The length of the Gh4CL genes is indicated by the scale line at the bottom. b Motif analysis. The motif analysis was performed by the MEME suite. Twenty motifs were detected and are displayed in color coded boxes. The length of proteins can be estimated using the scale line at the bottom.



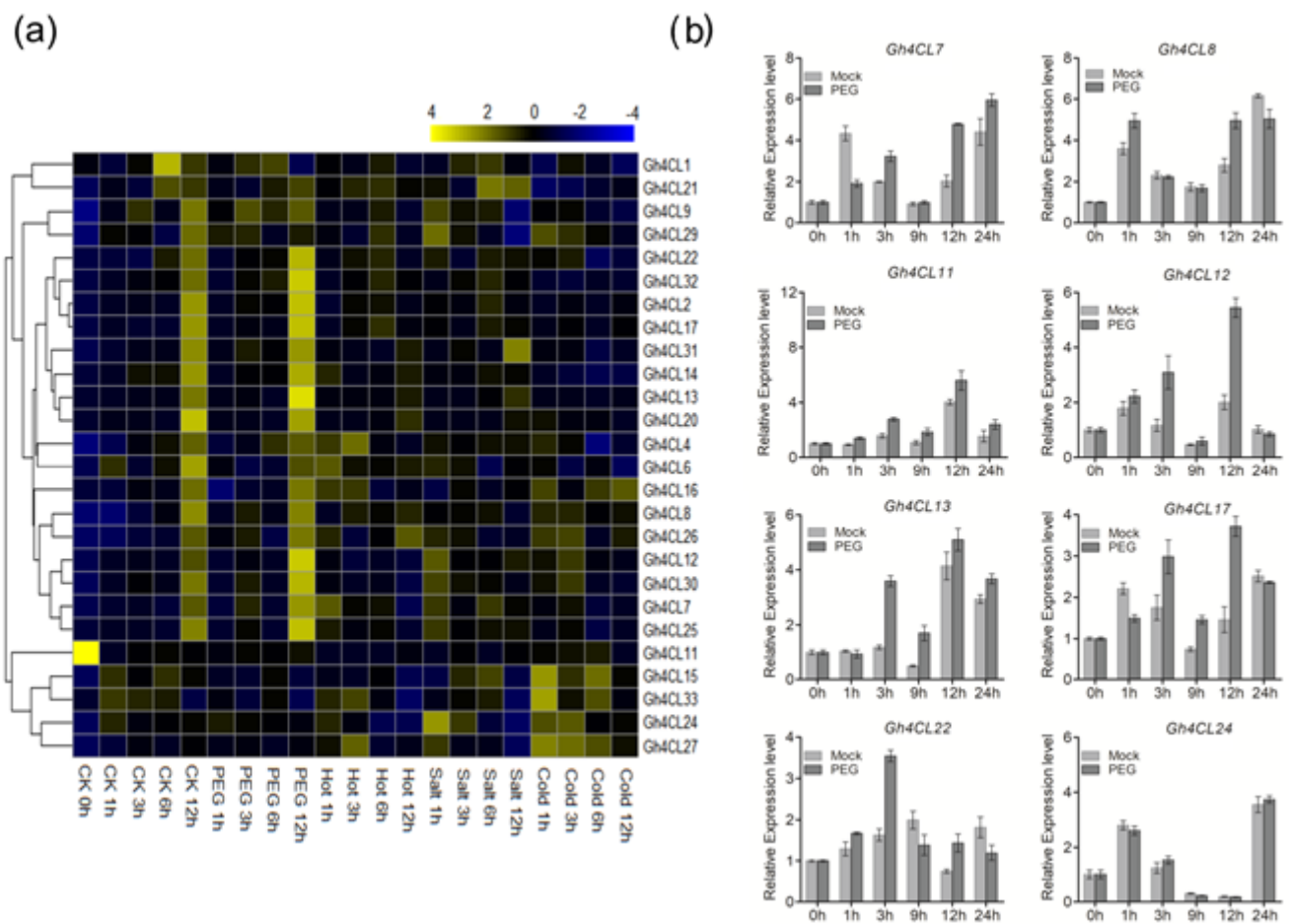
**Figure 3**

Promoter cis-element analysis of the Gh4CL genes. The numbers of different cis-elements are presented in the form of bar graphs.



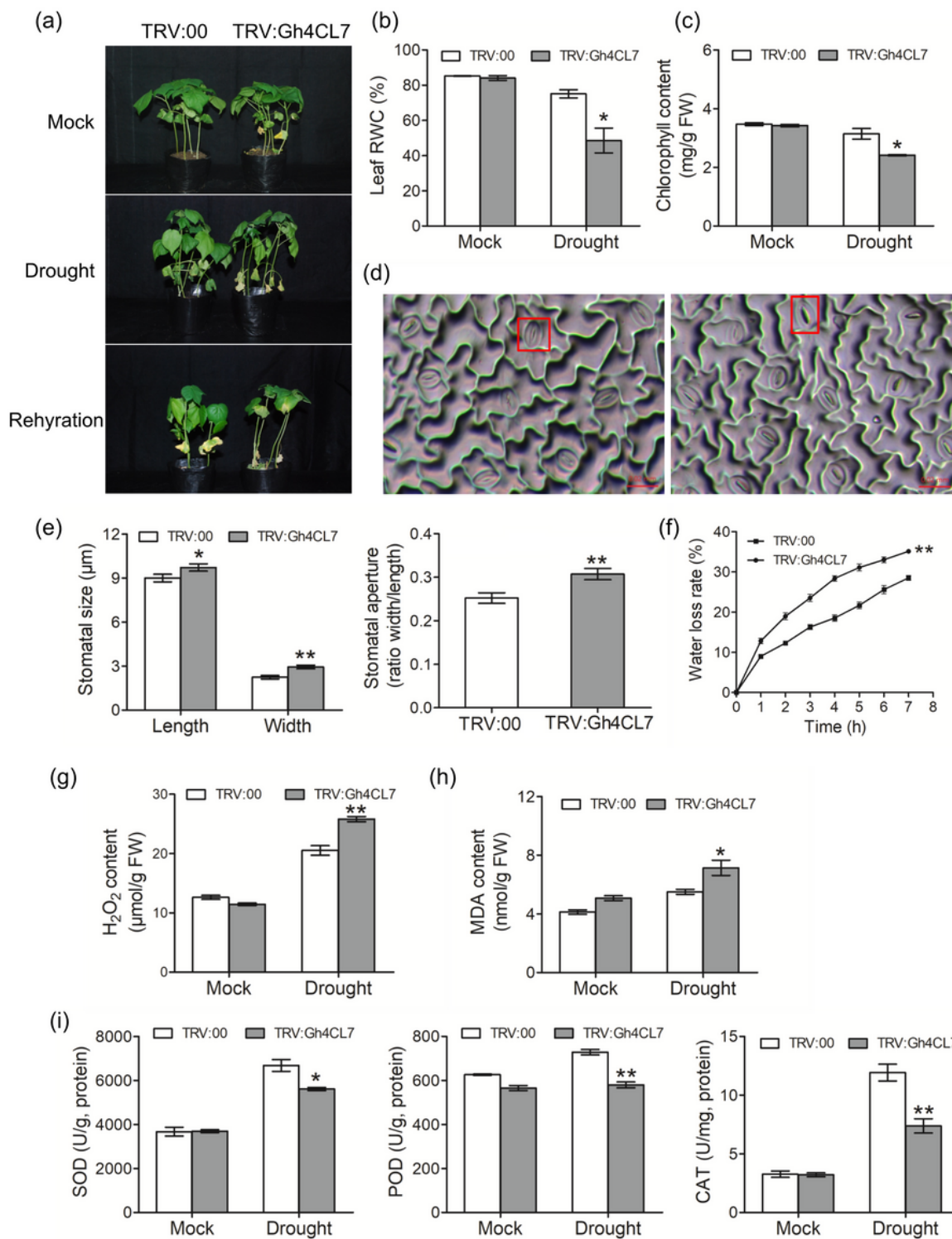
**Figure 4**

Expression pattern of the Gh4CL genes in different tissues of *G. hirsutum* based on the RNA-seq data generated by Zhang et al. (2015). The heatmap was generated by using the pheatmap package.



**Figure 5**

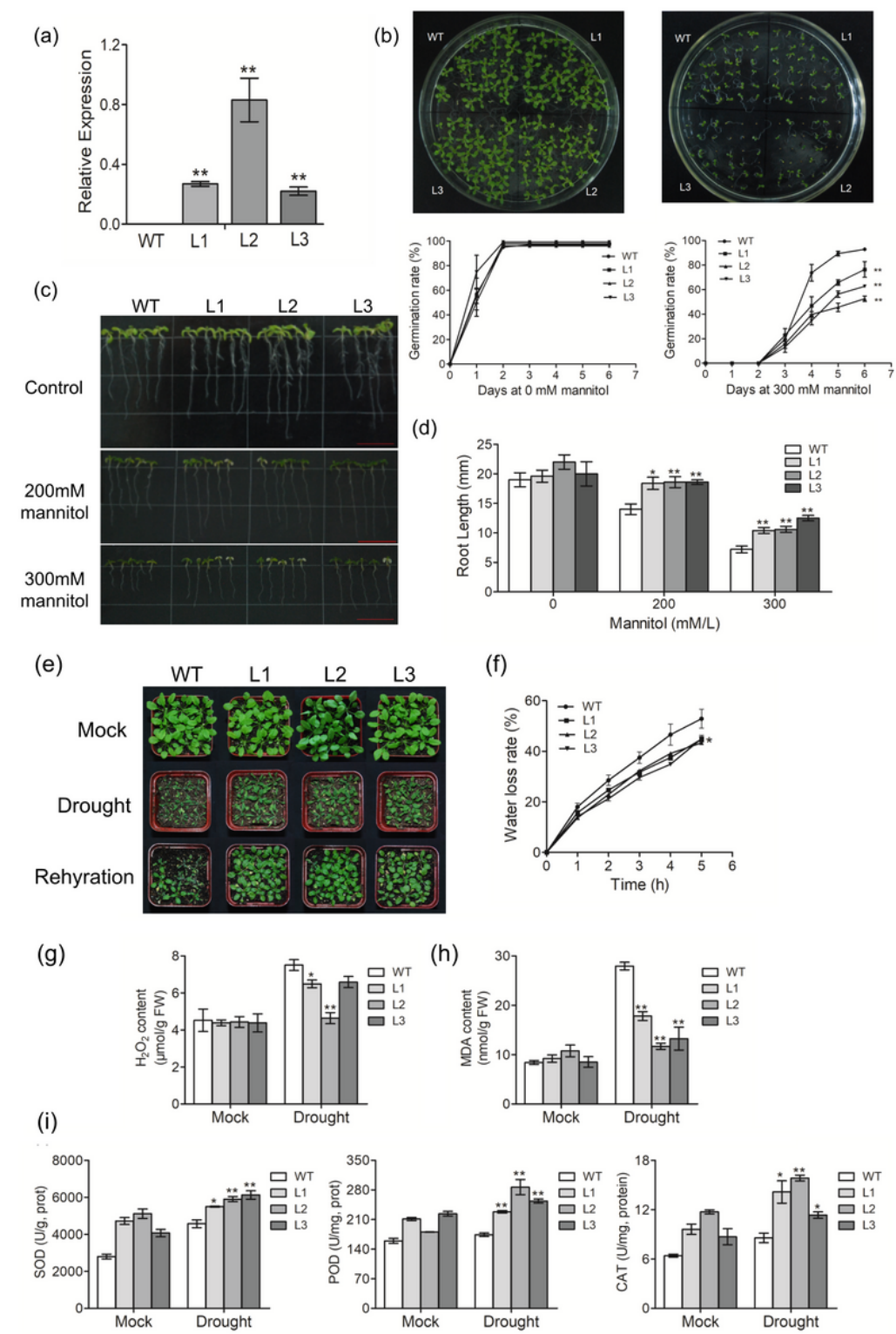
Expression profile of the Gh4CL genes in response to different abiotic stresses. a Heatmap showing the relative expression levels of each Gh4CL gene in each treatment showing at the bottom of the image. The results were based on the RNA-seq data generated by Zhang. b The relative expression level of the selected Gh4CL genes under 20% PEG stress based on qRT-PCR. The GhUBQ7 gene was used as internal control. Error bars represents the standard deviation calculated from three biological replicates.



**Figure 6**

Drought tolerance analysis of the Gh4CL7-silencing cotton plants. a Representative phenotypes of the TRV:00 (CK) and TRV:Gh4CL7 (VIGS) plants after three weeks of drought treatment. b-c Leaf RWC and chlorophyll contents of CK and VIGS plants under drought stress. d-e Comparison of stomata in the CK and VIGS plants. f Water loss rate of detached leaves from the CK and VIGS plants. g-h Comparison of hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>) and malondialdehyde (MDA) contents in the CK and VIGS plants. i Activities of

antioxidative enzymes in the CK and VIGS plants. Mock, normal growth conditions; drought, three weeks of water deficit conditions. Statistical significance was determined by the student test. \* and \*\* represent significant at  $P < 0.05$  and  $P < 0.01$ , respectively.

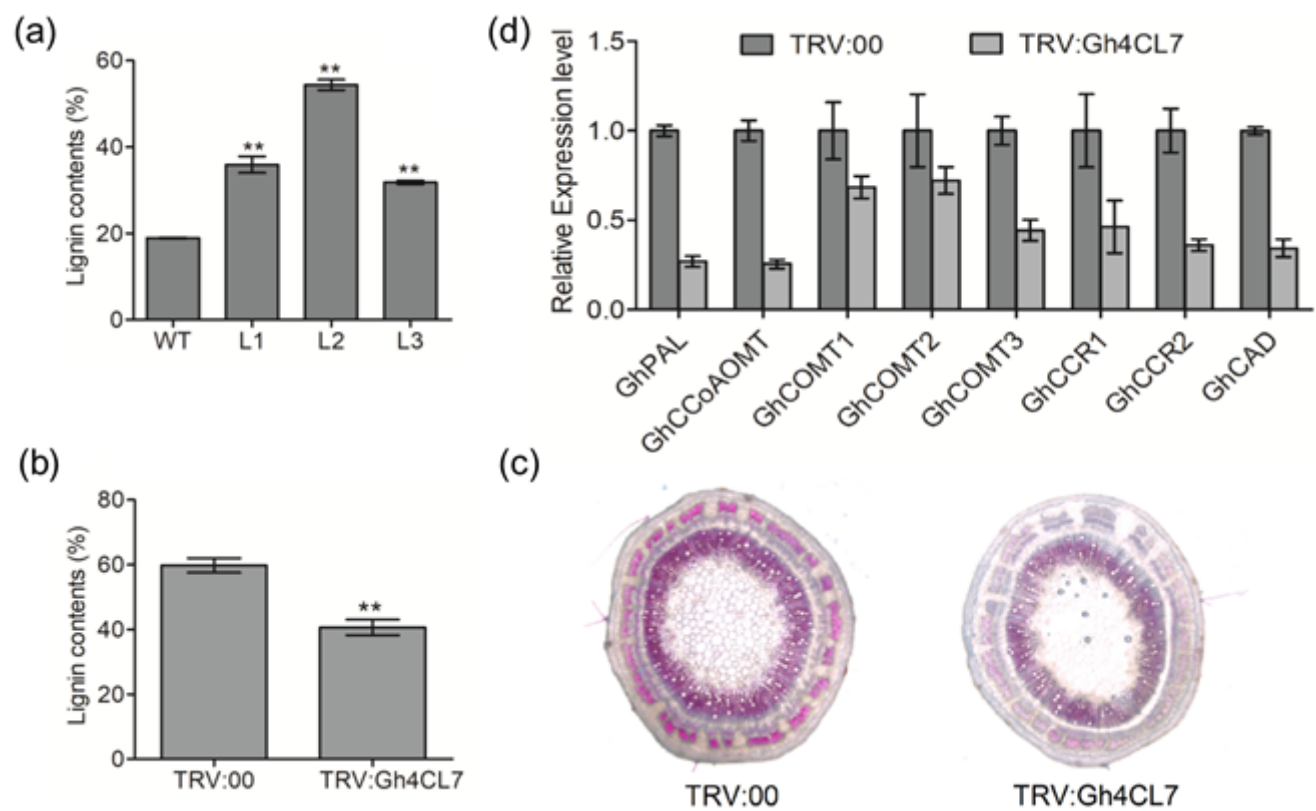


**Figure 7**

Drought tolerance analysis of the Gh4CL7-overexpressing Arabidopsis plants. a The relative expression level of Gh4CL7 in three independent Gh4CL7-OE Arabidopsis plants. b Germination rate of the Gh4CL7-



OE Arabidopsis seeds on 1/2 MS supplemented with 0 and 300mM mannitol. c-d Root elongation of the 6-days-old Gh4CL7-OE Arabidopsis seedlings on 1/2 MS supplemented with 0, 200, and 300mM mannitol. Bar = 1 cm. e Phenotypic assay of the Gh4CL7-OE and WT Arabidopsis plants during drought treatment. Ten-days-old seedlings were transplanted to soil and regularly watered for 2 weeks. For drought treatment, irrigation was terminated for two weeks. For the rehydration treatment, the plants were re-watered 3 days after the drought treatment. f Water loss rate of detached leaves from WT and Gh4CL7-OE Arabidopsis plants. g-i The contents of H<sub>2</sub>O<sub>2</sub> and MDA and antioxidative enzyme activity in Gh4CL7-OE and WT plants under normal and water deficit conditions. Data were represented as the mean  $\pm$  SE of three biological replicates. Statistical significance was determined by the student test. \* and \*\* represent significant at  $P < 0.05$  and  $P < 0.01$ , respectively.



**Figure 8**

Analyses of lignin contents and expression of the lignin biosynthesis related genes. a-b Comparison of lignin contents in Gh4CL7-OE Arabidopsis plants (a) and in Gh4CL7-silencing cotton plants (b). c Stem sections from the TRV:Gh4CL7 and TRV:00 cotton plants stained with phloroglucinol-HCl. d The relative expression levels of the lignin biosynthesis related genes. Data were represented as the mean  $\pm$  SE of three biological replicates; asterisks indicate levels of significance based on t-test (\*  $P < 0.05$ , \*\*  $P < 0.01$ ).

## Supplementary Files

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

- [Additionalfile1.doc](#)
- [Additionalfile2.doc](#)