Genome-wide Analysis of the CalS gene family in cotton reveals their potential roles in fiber development and responses to stress

Jiajia Feng [1, 2], **Yi Chen** [1], **Xianghui Xiao** [2], **Yunfang Qu** [1], **Pengtao Li** [2], **Quanwei Lu** [Corresp., 1, 2], **Jinling Huang** [Corresp. 1]

[1] College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi, China

[2] School of Biotechnology and Food Engineering, Anyang Institute of Technology, Anyang, Henan, China

Corresponding Authors: Quanwei Lu, Jinling Huang
Email address: daweianyang@163.com, huangjlsxau@163.com

Callose deposition occurs during plant growth and development, as well as when plants are under biotic and abiotic stress. *Callose synthase* is a key enzyme for the synthesis of callose. In this study, 27, 28, 16, and 15 callose synthase family members were identified in *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium raimondii*, and *Gossypium arboreum* using the sequence of *Arabidopsis callose synthase*. The *CalSs* were divided into five groups by phylogenetic, gene structure, and conservative motif analysis. The conserved motifs and gene structures of *CalSs* in each group were highly similar. Based on the analysis of cis-acting elements, it is inferred that *GhCalSs* were regulated by abiotic stress. WGD/Segmental duplication promoted the amplification of the CalS gene in cotton, and purification selection had an important function in the CalS family. The transcriptome data and qRT-PCR under cold, heat, salt, and PEG treatments showed that *GhCalSs* were involved in abiotic stress. The expression patterns of *GhCalSs* were different in various tissues. We predicted that *GhCalS4*, which was highly expressed in fibers, had an important effect on fiber elongation. Hence, these results help us understand the role of *GhCalSs* in fiber development and stress response.

Genome-wide Analysis of the CalS gene family in cotton reveals their potential roles in fiber development and responses to stress

Jiajia Feng[1,2], Yi Chen[1], Xianghui Xiao[2], Yunfang Qu[1], Pengtao Li[2], Quanwei Lu[1,2*], Jinling Huang [1*]
[1] College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi, China
[2] School of Biotechnology and Food Engineering, Anyang Institute of Technology, Anyang, Henan, China

* Co-Corresponding Author:
Quanwei Lu
Anyang, Henan, 455000, China
Email address: daweianyang@163.com
Jinling Huang
Taigu, Shanxi, 030600, China
Email address: huangjlsxau@163.com

Abstract

Callose deposition occurs during plant growth and development, as well as when plants are under biotic and abiotic stress. *Callose synthase* is a key enzyme for the synthesis of callose. In this study, 27, 28, 16, and 15 callose synthase family members were identified in *Gossypium hirsutum*, *Gossypium barbadense*, *Gossypium raimondii*, and *Gossypium arboreum* using the sequence of *Arabidopsis callose synthase*. The *CalSs* were divided into five groups by phylogenetic, gene structure, and conservative motif analysis. The conserved motifs and gene structures of *CalSs* in each group were highly similar. Based on the analysis of cis-acting elements, it is inferred that *GhCalSs* were regulated by abiotic stress. WGD/Segmental duplication promoted the amplification of the CalS gene in cotton, and purification selection had an important function in the CalS family. The transcriptome data and qRT-PCR under cold, heat, salt, and PEG treatments showed that *GhCalSs* were involved in abiotic stress. The expression patterns of *GhCalSs* were different in various tissues. We predicted that *GhCalS4*, which was highly expressed in fibers, had an important effect on fiber elongation. Hence, these results help us understand the role of *GhCalSs* in fiber development and stress response.

Keywords. Callose synthase, Synteny, Gene family, Gossypium, various stresses

Introduction

Callose is a linear homopolymer composed of *β*-1, 3-linked glucose, widely found in higher plants as an important part of specialized cell walls or cell wall-associated structures (Chen & Kim, 2009). However, Callose synthesis mainly depends on *callose synthase* (*CalS*) (Hong, Delauney & Verma, 2001). In *Arabidopsis thaliana*, callose activity was significantly increased and deposited after overexpression of the *AtCalS12* gene (Ellinger et al., 2013). In *rice*, callose

deposition in the plasmodesmata of *crr1* (*AtCalS10* homologous gene) mutants was reduced (Song et al., 2016).

Callose can regulate the transport of plasmodesmata and phloem, and affect plant development and response to biotic stress by controlling *callose synthases*. For example, the *AtCalS5* could maintain the normal formation of callose walls during pollen development (Shi et al., 2015). Meanwhile, *CalS* was regulated by various signaling pathways, and different biological regulatory processes involve hormones, transcription factors. For instance, external application of ABA can increase the activity of callose synthase of rice, promoting callose deposition, and thus increasing its resistance to BPH insects (Liu et al., 2017). The expression of *CalSs* were also affected by developmental and stress conditions, such as pollen development (Toller et al., 2008; Xie et al., 2010; Huang et al., 2009; Shi et al., 2015), low-temperature stimulation (Fromm et al., 2013), mechanical wounding (Cui & Lee 2016; Jacobs et al., 2003; Xie et al., 2011), fungal diseases (Dong et al., 2008; Oide et al., 2013; Ellinger et al., 2014; Blümke et al., 2013), bacterial diseases (Granato et al., 2019; Enrique et al., 2011), and insect diseases (Ahmad et al., 2011; Koh et al., 2012; Sun et al., 2018; Yao et al., 2020).

Due to the importance of callose, the callose synthase gene family has been identified in various plants. To date, it has been reported as 12 *AtCalSs* in *Arabidopsis thaliana* (Hong, Delauney & Verma, 2001), 12 *CsCalSs* in *Citrus sinensis* (Granato et al., 2019), 32 *BnCalSs* in *Brassica napus* (Liu et al., 2018), 8 *VvCalSs* in *Vitis vinifera* (Yu et al., 2016),7 *HvCalSs* in *Hordeum vulgare* (Schober et al., 2009), 15 *BraCalSs* in *Chinese cabbage* (Pu et al., 2019). Generally, according to the evolutionary analysis of *CalSs* in the above species, the CalS family can be divided into 3 or 4 main groups (Chen & Kim, 2009; Liu et al., 2018).

Cotton is an important economic crop in China which yield is affected by biotic and abiotic stresses, producing prevalent fibers for textile industry (Wang et al., 2020; Fang et al., 2014). Some researches have reported that cotton fiber elongation was related to callose deposition which may be involved in the closing of plasmodesmata, then promoted the fiber length (Ruan et al., 2004). It is possible that callose affects fiber elongation by controlling *CalS*. In light of the above, the *CalS* may play a significant role in cotton in responsing various stress and promoting fiber elongation. A callose synthase gene, *CFL1* was identified (Cui et al., 2001), however, the callose synthase gene family members, phylogenetic relationships and expression patterns in cotton are still unclear.

In the current study, we identified callose synthase genes in two cultivated allotetraploids cotton, *G. hirsutum* and *G. barbadense*, and their two putative genome donors, *G. raimondii* and *G. arboreum*, then discussed their phylogenetic relationships, conserved domains, gene structures, synteny, and cis-acting elements. We also focused on the expression patterns of *GhCalSs* in various tissues and their expression under abiotic stress. These findings provide a solid foundation for further study of the roles of *CalSs* in cotton fiber development and stress responses.

Materials & Methods

Plant Materials and Treatments.

Upland cotton TM-1, planted in Anyang Institute of Technology, was subjected to salt stress (350 mM NaCl) and drought stress (12% PEG6000) when the seedling reached two weeks. The leaves were collected 0h, 1h, 3h, 6h, 12h, and 24h after treatment. CCRI45 and MBI7747 were planted on farms managed by Cotton Research of Chinese Academy of Agricultural Sciences in Anyang. Cotton fibers were collected at 5, 10, 15, 20, 25 developmental time-points post-anthesis (DPA). All samples were stored at -80℃.

Identification of CalS family members in *Gossypium spp.*

The genome sequences and annotated files of *G. hirsutum* (Hu et al., 2019), *G. barbadense* (Hu et al., 2019), *G. raimondii* (Paterson et al., 2012), and *G. arboretum* (Du et al., 2018) were downloaded from Cottongen (https://www.cottongen.org/) (Yu et al., 2014). Both blast and HMMER were used to identify the CalS sequences. The 1, 3-beta-glucan synthase (PF02364) and FKS1_DOM1 domain (PF14288) from the Pfam database (http://pfam.xfam.org/) were searched by the HMMSearch program in TBTools to determine the presumed protein sequence (Chen et al., 2020). Besides, 12 *Arabidopsis CalSs* (Hong, Delauney & Verma, 2001) were used as queries sequences to identify family members using the Blastp program of TBtools (Chen et al., 2020). The protein sequences without above two domains were rejected and the domain which incomplete were also deleted. Finally, the final sequences were calculated by using ExPASy (https://www.expasy.org/) to calculate the theoretical isoelectric points (pI) and molecular weights (MW) and using the CELLO (http://cello.life.nctu.edu.tw/) for subcellular localization prediction (Yu et al., 2006).

Phylogenetic tree construction, gene structure, and motif analysis.

The phylogenetic tree among four Gossypium species and *Arabidopsis thaliana* was constructed by MEGA7(Kumar et al., 2016). It was constructed by the neighbor-joining (NJ) method, with 1000 bootstrap replicates, then was drawn by using EvolView (He et al., 2016). TBtools was used to extract the location information of *CalSs* and visualize the gene structure. MEME (https://meme-suite.org/meme/tools/meme) was used to identify the conservative motif with the parameter set to the maximum number of motifs: 20.

Chromosome location and synteny analysis for *CalSs*.

The locations of *CalSs* on chromosomes were shown by TBtools using four cotton species genomic annotation files (Chen et al., 2020). MCScanX was used to analyze the collinearity of the *CalSs*, that is, using CalS protein sequences to analyze the orthologous and paralogous gene pairs (Chen et al., 2020). Collinear gene pairs were visualized by using the circos (Chen et al., 2020). To investigate the selection pressure between homologous genes, we calculated the nonsynonymous substitutions rate (Ka) and synonymous substitutions rate (Ks) of homologous genes by KaKs_Calculator (Wang et al., 2010).

Analysis of Cis-acting element in promoters and functional enrichment analysis.

The 2000bp sequence upstream of the translation initiation codon ATG of CalS gene was selected as promoter. The cis-acting elements contained in the promoter region of the *CalSs* were predicted using the PlantCare website (Lescot et al., 2002). For functional enrichment analysis, the *GhCalSs* were submitted to the TBtools for GO annotation, and then use TBtools to visualize the results (Chen et al., 2020).

GhCalSs expression pattern under different tissues and abiotic stresses.

In order to analyze the expression of *GhCalSs* in different tissues and under stress, we downloaded 11 tissues (bract, pental, torus, root, leaf, stem, pistil, sepal, anther, ovule, fiber) and abiotic stress treatment (cold, heat, drought, salt) data from Cotton Omics Database (http://cotton.zju.edu.cn) (accession number: PRJNA490626) (Hu et al., 2019). *GhCalSs* with FPKM > 1 were considered as expressed genes. The expression patterns of the *GhCalSs* were visualized by ComplexHeatmap (Gu et al., 2016) based on the value of $\log_2$(FPKM+1).

RNA isolation and qRT-PCR analysis.

FastPure Plant Total RNA Isolation Kit (RC401, Vazyme) was used to extract RNA, and then we used 1μg to synthesize cDNA (HiScript III 1st Strand cDNA Synthesis Kit, R312 Vazyme). ChamQ Universal SYBR qRT-PCR Master Mix (Q711, Vazyme) was used for qPCR in ABI 7500 Fast Real-time PCR System (Applied Biosystems, USA). Gene-specific primers for qRT-PCR were designed by using prime-blast in NCBI, with melting temperatures of 55-60°C, product lengths of 101-221 bp, primer length of 18-25 bp (Table S1). For qRT-PCR, the reaction contains 10μL 2x ChamQ Universal SYBR qPCR Master Mix, 0.4μL of each primer, 3μL template, and ddH_2O to make up the total 20μL volume. Then it was carried out in the following condition: one cycles of 95°C for 30 s, 40 cycles of 95°C for 10 s and 60°C for 30 s. Each experiment was repeated three times, and two of the completed data were selected for drawing. Expression of all genes were calculated using a $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001).

Results

Identification and characterization of _CalSs_ in _Gossypium spp_.

Through the analysis of the CalS protein sequences in *Arabidopsis thaliana*, we found that all the protein sequences contain 1, 3-*β*-glucan synthase (PF02364) and FKS1_DOM1 domain (PF14288), total of 27 members of the CalS gene family in *G. hirsutum*, 28 in *G. barbadense*, 15 in *G. raimondii*, and 16 in *G. arboretum* were identified, all of which were named according to their chromosomal locations. The properties of *CalSs* in cotton were further analyzed (Table S2). The protein sequence length of *CalSs* ranged from 1494 to 1979 amino acids, with an average MW of 212.88 kDa, and shared high similarity to the *Arabidopsis thaliana* CalS proteins (Hong, Delauney & Verma, 2001). The isoelectric point (PI) values of the above genes were all greater than 7, indicating that *CalSs* in cotton were alkaline, which was the same as the biochemical properties of the *CalSs* in *Chinese cabbage* and *Brassica*(Pu et al., 2019; Liu et al., 2018). The *CalSs* were most likely localized in the plasma membrane, as predicted in *Arabidopsis thaliana* and *Chinese cabbage*(Pu et al., 2019; Zavaliev et al., 2011).

Classification and phylogenetic analysis of the cotton _CalSs_.

In order to investigate the evolutionary relationships of the *CalSs* in the four cotton species and its relationship with *Arabidopsis thaliana*, a phylogenetic tree was constructed using the protein sequences of *CalSs* (Fig. 1). Based on the phylogenetic tree of this study, the *CalSs* were divided into five groups. The distribution of *CalSs* in each group was shown in Table S3. The members of Group A were homologous to *AtCalS11/AtCalS12*, the members of Group B were homologous to *AtCalS9/AtCalS10*, the members of Group C were homologous to

AtCalS6/AtCalS7, the members of Group D were homologous to *AtCalS8*, and the members of Group E were homologous to *AtCalS1-5*. There were *Arabidopsis* genes homologous to cotton in each group, further indicating that the cotton *CalSs* and the *Arabidopsis thaliana CalSs* were close in evolutionary, which was consistent with the evolutionary relationship between *Arabidopsis* and cotton. It is observed that most of the *CalSs* derived from At-subgenome of two cultivated allotetraploids cotton stayed close together with the CalS gene of *G. arboretum*, and the *CalSs* of Dt-subgenome stayed close together with the CalS gene of *G. raimondii*, which was consistent with the hypothesis that the allotetraploid cotton species were produced by the recombination of two diploid cotton species (Liu et al., 2015). Phylogenetic tree analysis suggested that the CalS homologous gene in cotton may have similar functions.

Gene structure and amino acid motif analysis of the *CalSs*.

The diversity of gene structure and differences in conserved motifs are the manifestations of the evolution of multigene families (Magwanga et al., 2018). The distribution of exon/intron regions of *CalSs* was analyzed to realize the diversity of gene structure (Fig. 2). The number of *CalSs* exons varied from 1-51, and most *CalSs* had more than 35 exons (57/86, 66.2%). Clearly, these *CalSs* were divided into an exon-poor group (<7 exons, group A) and other exon-rich groups (>37 exons, group B-E) (Fig2, Table S2). The exons of *CalSs* had high similarity in the same group, and the number of exons in group B, group D, and group E were the same (Fig2, Table S2).

The motif is a conserved region in the sequence (Magwanga et al., 2018). We identified 20 possible motifs using MEME (Fig. 2). Interestingly, all *CalSs* except *GbCalS2/14/15* and *GhCalS1* contained motif1-20 and were arranged in the sequence of motif15-9-8-13-12-11-7-20-14-16-3-2-1-6-19-5-17-4-18-10. The distribution of *CalSs* were slightly different among different groups, and only the number and arrangement position were different. The number and arrangement of motifs in the same group of *CalSs* were more similar than those in other groups.

Chromosomal location, gene duplication, and syntenic analysis of the *CalSs* in *Gossypium spp*.

Based on the sequencing and annotated information of the four cotton genomes, the chromosome length and the distribution of genes on chromosome could be analyzed (Fig. 3). The distribution of *CalSs* in the two heterotetraploid cotton species chromosomes was highly similar. For example, *CalSs* had the same number and distribution on chromosomes A03, A04, A05, A08, A11. In *G. hirsutum*, 27 *GhCalSs* were distributed on 15 chromosomes, including 13 *GhCalSs* in At-subgenome and 14 *GhCalSs* in Dt-subgenome. In *G. barbadense*, 28 *GbCalSs* were distributed on 16 chromosomes, including 14 *GbCalSs* in At-subgenome and 14 *GbCalSs* in Dt-subgenome. In *G. arboreum*, 16 *GaCalSs* were distributed on 8 chromosomes and a scaffold. In *G. raimondii*, 15 *GrCalSs* were distributed on 8 chromosomes. Most *CalSs* occured at the upper or lower arms of chromosomes. D08 and D10 chromosomes both had the largest number of *CalSs* in the two allotetraploid cotton. Obviously, chromosome length was not positively correlated with the distribution number of *CalSs* on chromosomes.

Gene duplication is the basis for the functional differentiation of homologous genes, the main reason for the generation of new functional genes(Conant & Wolfe, 2008). In order to

explain the gene replication events of *CalSs* in cotton, we identified 15, 14 paralogous gene pairs in *G. hirsutum*, *G. barbadense* respectively, and 1 pair in *G. arboreum*. But there was no paralogous gene pair in *G. raimondii* (Table S4). *GhCalS21/22*, *GbCalS1/2* as well as *GbCalS22/23* were tandem duplication. In the four cotton species, the duplication events of the *CalSs* were WGD/ Segmental, Tandem Duplicates, Dispersed, and proximal duplication, and the main expansion mechanism was WGD/Segmental.

In order to illustrate the collinearity of CalS genes, we analyzed the orthologous and paralogous gene pairs (Fig. 4, Table S4). There were 31 CalS orthologous gene pairs among *G. arboretum* and two allotetraploid cotton species, including 15 pairs between with At-subgenome of *G. hirsutum* and 16 pairs between with At-subgenome of *G. barbadense*. There were 17 CalS orthologous gene pairs among *G. raimondii* and two allotetraploid cotton species, including 9 pairs between with Dt-subgenome of *G. hirsutum* and 8 pairs between with Dt-subgenome of *G. barbadense*. Meanwile, Ka/Ks of CalS homologous pairs were calculated to further understand the adaptation of the CDS region of *CalSs* (Fig. 4, Table S5). Most of the homologous gene pairs Ka/Ks <1, and about 94.6% gene pairs had a Ka/Ks ratio less than 0.5, which meant that almost all gene pairs underwent purification selection. Only Ka/Ks >1 of *GaCalS2 / GbCalS1* indicated that this was a positive selection for beneficial mutations.

Analysis of Cis-acting elements in promoter.

Transcription factors can be combined with cis-elements in the promoter region to regulate gene transcription. Investigation of upstream regulatory sequence can help us to well understand the regulation mechanism and also supportive to estimate the potential function of the gene. (Fig. 5, Table S6). Given the effect of plant hormones in abiotic stress, we focused on plant hormone responsive elements in promoter regions. ABA-(ABRE), auxin-(AuxRR-core, TGA-element), Gibberellin- (GARE-motif, P-box, TATC-box), MeJA-(CGTCA-motif, TGACG-motif), SA-(TCA-element) responsive element were found in the promoters of 18, 6, 14, 18, 12 *GhCalSs*. All *GhCalSs* contained hormone response elements except the *GhCalSs* in Group D and *GhCalS5* in Group C. More than half of the *GhCalSs* contained ABA/GA/MeJA responsive element. Auxin responsive element only in *GhCalS6/14/20/21/22/23*. Meanwhile, we also paid attention to elements related to stress. Low-temperature- (LTR), wound- (WUN-motif), drought-(MBS), stress- (TC-rich repeats), anaerobic induced response element (ARE), anoxic specific inducibility element (GC-motif) were found in the promoters of 11, 2, 15, 10, 23, 4 *GhCalSs*. Wound-responsive elements only in *GhCalS5* and *GhCalS18*. Anoxic specific inducibility element only in *GhCalS11/15/17/25*. In addition, these results suggested that *CalSs* might regulated by hormone and abiotic stresses.

Expression patterns of the *GhCalSs* under abiotic stresses.

Previous studies have reported that the *CalSs* respond to abiotic stresses (Cui and Lee 2016; Jacobs AK; Fromm et al., 2013). To understand the response of *GhCalSs*, we used public RNA-seq data of TM-1 treated with cold, hot, NaCl, and PEG to observe the expression pattern of *GhCalSs* (Fig. 6). Interestingly, all expressed *GhCalSs* were induced by different abiotic stress, and the expression patterns were different. The expression of *GhCalS3* was significantly up-regulated under cold, hot, NaCl, and PEG. The expression patterns of *GhCalSs* in the same group

were slightly consistent, such as *GhCalS3* and *GhCalS6*, *GhCalS2* and *GhCalS16*. In order to verify the results obtained by the above transcriptome, cotton seedlings were treated with PEG and NaCl, and then the *GhCalS2/3/6/9/16* were selected for qRT-PCR (Fig. 7). The expression of *GhCalS3* and *GhCalS6* in Group A were up-regulated within 24 hours under PEG treatment and reached the peak at 24 hours. The expression of *GhCalS2*, *GhCalS9*, *GhCalS16* were up-regulated at first and then down-regulated and last up-regulated after PEG treatment. After NaCl treatment, there was no consistent trend of gene expression. *GhCalS3* was significantly induced by NaCl and significantly up-regulated at 3 h. Both *GhCalS2* and *GhCalS16* of Group B were down-regulated within 24 hours. The expression of *GhCalS6* and *GhCalS9* reached a peak at 12 h. These findings indicated that the expression patterns of several *GhCalSs* were changed after treatment, which proved that *GhCalSs* increased adaptability to abiotic stress (Fig. 6).

Enrichment analysis of the *GhCalSs*.

In order to further understand the function of *GHCalSs*, we carried out functional enrichment annotation of gene ontology (GO). The results improve our accurate understanding of gene function, including many significantly enriched terms (Fig8, S7 Table). The GO-BP enrichment results showed 27 terms such as beta-glucan metabolic process (GO:0051273), cellular carbohydrate biosynthetic process (GO:0034637), (1->3)-beta-D-glucan biosynthetic process (GO:0006075), cellular macromolecule biosynthetic process (GO:0034645). The GO-CC enrichment results discovered 11 terms such as transferase complex (GO:1990234), plasma membrane protein complex (GO:0098797), 1,3-beta-D-glucan synthase complex (GO:0000148), catalytic complex (GO:1902494), etc. The CC terms enriched by GO were consistent with the subcellular localization of *GHCalSs*. GO-MF enrichment exposed 8 terms, including 1,3-beta-D-glucan synthase activity (GO:0003843), UDP-glucosyltransferase activity (GO:0035251), catalytic activity (GO:0003824), hexosyltransferase activity (GO:0016758). In short, the GO enrichment results confirmed the function of the *GhCalSs* in many biological processes, which were associated with 1,3-β-D-glucan synthetic activity, hydrolyzase activity, and membrane parts.

***GhCalSs* expression patterns in various tissues and their role in fiber development.**

We used transcriptome data from different tissues *GhCalSs* to gain insight the tissue-specific expression patterns of cotton. For instance, *GhCalSs* were expressed in various tissues, and some of them were highly expressed. Some of *GhCalSs* were expressed in one or more tissues (*GhCalS2, 3, 4, 6, 8, 9, 15, 16, 19, 20, 21*). However, the expression of a few genes (*GhCalS1, 5, 7, 10, 11, 12, 13, 14, 17, 18, 22, 23, 24, 25, 26, 27*) did not show any expression in any tissues.

In order to determine the effect of *GhCalSs* in cotton fiber development, we focused on the expression of *GhCalSs* in different fiber development stages of two samples, MBI7747 and CCRI45, with different lengths and strengths (Lu et al., 2017) (Fig. 9). The expression of *GhCalS4* was the highest in TM-1, MBI7747, CCRI45 fiber tissue, so it was speculated that *GhCalS4* had an important function in cotton fiber development. In order to further determine its function in fiber development, qRT-PCR was used to analyze the *GhCalS4* expression differences in two samples (Fig. 9). The results showed that the expression level of *GhCalS4* in the two samples gradually increased from 5 DPA to 25 DPA, which was consistent with the

transcriptome data of TM-1 used above. The expression of *GhCalS4* in CCRI45 was higher than in MBI7747 at 5DPA, 10DPA, 15DPA but was significantly lower than that of MBI7747 at 25DPA (Lu et al., 2017). Thus, *GhCalS4* may be involved in cotton fiber elongation.

Discussion

Callose plays a vital role in plant growth, development, and resistance to various adverse factors (Piršelová & Matušíková 2012). The gene family of callose synthase has been identified in a variety of plants. In this study, we identified the *CalSs* in *G. hirsutum*, *G. barbadense*, *G. raimondii*, and *G. arboreum*, aiming to understand the role of CalS family in the cotton development.

A total of 86 *CalSs* were identified in four cotton species. They were divided into five groups based on evolutionary relationships. we divided *CalSs* (*AtCalS9-12* homologous) into two groups due to the large difference of the CDS number, and the other groups were the same as those in *Arabidopsis thaliana*. The *CalSs* number was 2:2:1:1 in group B/C/D, which was consistent with the evolutionary relationship among cotton species (Table S3). Compared with *Arabidopsis thaliana*, different percentages existed between subgroups. The percentages of Group A and Group E were significantly different from those of the corresponding *CalSs* in *Arabidopsis thaliana*, suggesting that these genes in cotton may have functional differences with homologous genes in *Arabidopsis thaliana* to a certain extent. These results will help to validate the function of cotton CalS homologous gene with *Arabidopsis thaliana*.

Tetraploid cotton species are formed by natural crossbreeding between *G. raimondii* and *G. arboretum* (Wendel et al., 2009). Thus, the four cotton species are closely related in evolution. In Figure 4, the direct homologous gene pairs of *CalSs* were all clustered in the same branch or group. Phylogenetic and direct homologous genes of CalS further indicated that the results of this study were consistent with the evolutionary view.

A large number of hormone-responsive elements were identified on *GhCalSs* promoters which may be involved in the regulation of *GhCalSs*. Salicylic acid (SA) was an endogenous signal molecule in plants (Loake & Grant, 2007). In *Arabidopsis thaliana*, the expressions of *AtCalS1/5/9/10/12* were up-regulated by exogenous SA. Abscisic acid (ABA) played an important part in coping with a variety of adverse factors, closely related to callose synthesis (Liu et al., 2017). During the dormancy of *Populus tomentosa* buds, short-day induced ABA biosynthesis, promoted the expression of *PtCalS1*, callose deposited at the plasmodesmata to form blockage, which prevented the growth signal molecules from entering the cell and kept the dormancy state of buds (Tylewicz et al., 2018). Jasmonic acid (JA) was also involved in callose regulation, and Methyl Jasmonate (MeJA) application promoted callose deposition in grape leaves. Inhibition of the expression of *Cationic peroxidase 3* (*OCP3*), a negative regulator of the JA pathway, increased callose deposition (Repka, Fischerová & Šilhárová, 2004). In conclusion, ABA, JA, and SA were involved in the regulation of callose deposition. *GhCalSs* promoters with ABA, SA and JA response elements were highly likely to be regulated by them in cotton. However, how CalS gene is regulated by these hormones in the face of biotic-abiotic stress or growth and development is not known, which needs to be further studied.

Callose deposition is one of a series of coping strategies in plants to abiotic stress. Low temperature stimulation of *maize* leaves increased callose content and reduced transport of photosynthate in phloem (Wu et al., 2018). In *Arabidopsis thaliana*, *AtCalS7*, *AtCalS8* and *AtCalS12* were associated with callose synthesis under the condition of wound (Jacobs et al., 2003; Cui and Lee 2016; Xie et al., 2011). In this study, public transcriptome data were used to analyze the responses of cotton leaves to cold, heat, salt and drought, and qRT-PCR was used to verify the results, which showed that *CalSs* were involved in abiotic stresses.

Callose deposits regulate material transport and control plant development. In this study, *GhCalS4* was highly expressed in fibers and differentially expressed in MBI7747 and CCRI45 fibers at each fiber development stage (5/10/15/20/25 DPA). It has been reported that callose deposition may be involved in the closure of plasmodesmata, and the closure of plasmodesmata has an important function in the elongation of cotton fibers (Ruan et al., 2004). In Island cotton, plasmodesmata remain open longer than in Upland cotton, allowing sucrose to be fed into fibroblasts, which eventually increase osmotic potential by hydrolysis to fructose and glucose. The more soluble sugar, K^+ accumulated, the higher the cell leavening pressure, which promoted the elongation of cotton fiber (Hu et al., 2019). MBI7747 is a chromosome segment substitution line（CSSL）with different genetic background constructed by crosses between the upland cotton CCRI45 as the recurrent parent and the Sea Island cotton Hai 1 with outstanding fiber quality through the combination of high-generation backcrossing and molecular marker-assisted selection. The fiber length and strength of MBI7747 are better than CCRI45. In CCRI45, the expression level of this gene at 5DPA, 10DPA and 15DPA were all higher than those of MBI7747 during fiber elongation, and it was speculated that the degree of callose deposition in MBI7747 was lower than that of CCRI45, which made more sucrose input into fiber cells to increase osmotic potential and promote fiber elongation. Thus, *GhCalS4* may be an introgression gene or there was difference in epigenetic regulation.

Conclusions

In this study, we identified 86 *CalSs* from *G. hirsutum*, *G. barbadense*, *G. raimondii*, and *G. arboreum* using conserved domains. Phylogeny, gene structure, motif, chromosome location and homologous genes were analyzed. It indicates that CalSs have been highly conserved during evolution by the analysis of *CalSs* structure, conversed motifs, and syntenic blocks. WGD/Segmental replication was the main driving force for the amplification of CalS family in cotton, and purification selection played an important role in the evolution of *CalSs*. In addition, the cis-acting elements of *GhCalSs* related to hormone regulation and development and its expression pattern in stress and tissue were also analyzed. CalS gene can be induced by abiotic stress. Some genes in Group A may have the important function in the development of cotton tissues. Furthermore, the expression difference of *GhCalS4* in fiber of different length and strength materials was analyzed. It was speculated that *GhCalS4* played a major role in fiber elongation. These findings could lay a foundation for further study on the role of CalS gene in stress response and fiber development.

References

Ahmad S, Veyrat N, Gordon-Weeks R, Zhang Y, Martin J, Smart L, Glauser G, Erb M, Flors V, Frey M, Ton J. 2011. Benzoxazinoid metabolites regulate innate immunity against aphids and fungi in maize. Plant Physiology **157 (1)**:317-327. DOI 10.1104/pp.111.180224

Blümke A, Somerville SC, Voigt CA. 2013. Transient expression of the Arabidopsis thaliana callose synthase PMR4 increases penetration resistance to powdery mildew in barley. Advances in Bioscience and Biotechnology **04 (08)**:810-813. DOI 10.4236/abb.2013.48106

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Molecular Plant **13 (8):**1194-1202. DOI 10.1016/j.molp.2020.06.009

Chen XY, Kim JY. 2009. Callose synthesis in higher plants. Plant Signaling and Behavior. **4 (6):**489-492. DOI 10.4161/psb.4.6.8359

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. Nature Reviews Genetics **9 (12):**938-950. DOI 10.1038/nrg2482

Cui W, Lee JY. 2016. Arabidopsis callose synthases CalS1/8 regulate plasmodesmal permeability during stress. Nature Plants **2 (5):**16034. DOI 10.1038/nplants.2016.34

Cui X, Shin H, Song C, Laosinchai W, Amano Y, Brown RM, Jr. 2001. A putative plant homolog of the yeast beta-1,3-glucan synthase subunit FKS1 from cotton (Gossypium hirsutum L.) fibers. Planta **213 (2):**223-230. DOI 10.1007/s004250000496

Dong X, Hong Z, Chatterjee J, Kim S, Verma DP. 2008. Expression of callose synthase genes and its connection with Npr1 signaling pathway during pathogen infection. Planta **229 (1):**87-98. DOI 10.1007/s00425-008-0812-3

Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M, Jia Y, Pan Z, Gong W, Liu Z, Zhu H, Ma L, Liu F, Yang D, Wang F, Fan W, Gong Q, Peng Z, Wang L, Wang X, Xu S, Shang H, Lu C, Zheng H, Huang S, Lin T, Zhu Y, Li F. 2018. Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. Nature Genetics **50 (6):**796-802. DOI 10.1038/s41588-018-0116-x

Ellinger D, Glockner A, Koch J, Naumann M, Sturtz V, Schutt K, Manisseri C, Somerville SC, Voigt CA. 2014. Interaction of the Arabidopsis GTPase RabA4c with its effector PMR4 results in complete penetration resistance to powdery mildew. Plant Cell **26 (7):**3185-3200. DOI 10.1105/tpc.114.127779

Ellinger D, Naumann M, Falter C, Zwikowics C, Jamrow T, Manisseri C, Somerville SC, Voigt CA. 2013. Elevated early callose deposition results in complete penetration resistance to powdery mildew in Arabidopsis. Plant Physiology **161 (3):**1433-1444. DOI 10.1104/pp.112.211011

Enrique R, Siciliano F, Favaro MA, Gerhardt N, Roeschlin R, Rigano L, Sendin L, Castagnaro A, Vojnov A, Marano MR. 2011. Novel demonstration of RNAi in citrus reveals importance of citrus callose synthase in defence against Xanthomonas citri subsp. citri. Plant Biotechnology Journal **9 (3):**394-407. DOI 10.1111/j.1467-7652.2010.00555.x

Fang L, Tian R, Li X, Chen J, Wang S, Wang P, Zhang T. 2014. Cotton fiber elongation network revealed by expression profiling of longer fiber lines introgressed with different

Gossypium barbadense chromosome segments. BMC Genomics 15:838. DOI https://doi.org/10.1186/1471-2164-15-838

Fromm J, Hajirezaei MR, Becker VK, Lautner S. 2013. Electrical signaling along the phloem and its physiological responses in the maize leaf. Frontiers in Plant Science **4:**239. DOI 10.3389/fpls.2013.00239

Granato LM, Galdeano DM, D'Alessandre NDR, Breton MC, Machado MA. 2019. Callose synthase family genes plays an important role in the Citrus defense response to Candidatus Liberibacter asiaticus. European Journal of Plant Pathology **155 (1)**:25-38. DOI 10.1007/s10658-019-01747-6

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics **32 (18):**2847-2849. DOI 10.1093/bioinformatics/btw313

He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. 2016. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Research **44 (W1):**W236-241. DOI 10.1093/nar/gkw370

Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, Baruch K, Fang D, Liu X, Ruan YL, Rahman MU, Han J, Wang K, Wang Q, Wu H, Mei G, Zang Y, Han Z, Xu C, Shen W, Yang D, Si Z, Dai F, Zou L, Huang F, Bai Y, Zhang Y, Brodt A, Ben-Hamo H, Zhu X, Zhou B, Guan X, Zhu S, Chen X, Zhang T. 2019. Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton. Nature Genetics **51 (4):**739-748. DOI 10.1038/s41588-019-0371-5

Huang L, Chen XY, Rim Y, Han X, Cho WK, Kim SW, Kim JY. 2009. Arabidopsis glucan synthase-like 10 functions in male gametogenesis. Journal of Plant Physiology **166 (4):**344-352. DOI 10.1016/j.jplph.2008.06.010

Jacobs AK, Lipka V, Burton RA, Panstruga R, Strizhov N, Schulze-Lefert P, Fincher GB. 2003. An Arabidopsis Callose Synthase, GSL5, Is Required for Wound and Papillary Callose Formation. Plant Cell **15 (11):**2503-2513. DOI 10.1105/tpc.016097

Koh EJ, Zhou L, Williams DS, Park J, Ding N, Duan YP, Kang BH. 2012. Callose deposition in the phloem plasmodesmata and inhibition of phloem transport in citrus leaves infected with "Candidatus Liberibacter asiaticus". Protoplasma **249 (3):**687-697. DOI 10.1007/s00709-011-0312-3

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution **33 (7):**1870-1874. DOI 10.1093/molbev/msw054

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

Liu F, Zou Z, Fernando WGD. 2018. Characterization of Callose Deposition and Analysis of the Callose Synthase Gene Family of Brassica napus in Response to Leptosphaeria maculans. International Journal of Molecular Sciences **19 (12).** DOI 10.3390/ijms19123769

Liu J, Du H, Ding X, Zhou Y, Xie P, Wu J. 2017. Mechanisms of callose deposition in rice regulated by exogenous abscisic acid and its involvement in rice resistance to Nilaparvata lugens Stal (Hemiptera: Delphacidae). Pest Management Science **73 (12):**2559-2568. DOI 10.1002/ps.4655

Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L, Zhou Y, Wang LJ, Guo WZ, Bai YL, Ruan JX, Shangguan XX, Mao YB, Shan CM, Jiang JP, Zhu YQ, Jin L, Kang H, Chen ST, He XL, Wang R, Wang YZ, Chen J, Wang LJ, Yu ST, Wang BY, Wei J, Song SC, Lu XY, Gao ZC, Gu WY, Deng X, Ma D, Wang S, Liang WH, Fang L, Cai CP, Zhu XF, Zhou BL, Jeffrey Chen Z, Xu SH, Zhang YG, Wang SY, Zhang TZ, Zhao GP, Chen XY. 2015. Gossypium barbadense genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Scientific Reports **5:**14139. DOI 10.1038/srep14139

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

Loake G, Grant M. 2007. Salicylic acid in plant defence--the players and protagonists. Current Opinion in Plant Biology **10 (5):**466-472. DOI 10.1016/j.pbi.2007.08.008

Lu Q, Shi Y, Xiao X, Li P, Gong J, Gong W, Liu A, Shang H, Li J, Ge Q, Song W, Li S, Zhang Z, Rashid MHO, Peng R, Yuan Y, Huang J. 2017. Transcriptome Analysis Suggests That Chromosome Introgression Fragments from Sea Island Cotton (Gossypium barbadense) Increase Fiber Strength in Upland Cotton (Gossypium hirsutum). G3-Genes Genomes Genetics **7 (10):**3469-3479. DOI 10.1534/g3.117.300108

Magwanga RO, Lu P, Kirungu JN, Cai X, Zhou Z, Wang X, Diouf L, Xu Y, Hou Y, Hu Y, Dong Q, Wang K, Liu F. 2018. Whole Genome Analysis of Cyclin Dependent Kinase (CDK) Gene Family in Cotton and Functional Evaluation of the Role of CDKF4 Gene in Drought and Salt Stress Tolerance in Plants. International Journal of Molecular Sciences **19 (9):**2625. DOI 10.3390/ijms19092625

Oide S, Bejai S, Staal J, Guan N, Kaliff M, Dixelius C. 2013. A novel role of PR2 in abscisic acid (ABA) mediated, pathogen-induced callose deposition in Arabidopsis thaliana. New Phytologist **200 (4):**1187-1199. DOI 10.1111/nph.12436

Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, Yoo MJ, Byers R, Chen W, Doron-Faigenboim A, Duke MV, Gong L, Grimwood J, Grover C, Grupp K, Hu G, Lee TH, Li J, Lin L, Liu T, Marler BS, Page JT, Roberts AW, Romanel E, Sanders WS, Szadkowski E, Tan X, Tang H, Xu C, Wang J, Wang Z, Zhang D, Zhang L, Ashrafi H, Bedon F, Bowers JE, Brubaker CL, Chee PW, Das S, Gingle AR, Haigler CH, Harker D, Hoffmann LV, Hovav R, Jones DC, Lemke C, Mansoor S, ur Rahman M, Rainville LN, Rambani A, Reddy UK, Rong JK, Saranga Y, Scheffler BE, Scheffler JA, Stelly DM, Triplett BA, Van Deynze A, Vaslin MF, Waghmare VN, Walford SA, Wright RJ, Zaki EA, Zhang T, Dennis ES, Mayer KF, Peterson DG, Rokhsar DS, Wang X, Schmutz J. 2012. Repeated polyploidization of

Gossypium genomes and the evolution of spinnable cotton fibres. Nature **492 (7429):**423-427. DOI 10.1038/nature11798

Piršelová B, Matušíková I. 2012. Callose: the plant cell wall polysaccharide with multiple biological functions. Acta Physiologiae Plantarum **35 (3):**635-644. DOI 10.1007/s11738-012-1103-y

Pu Y, Hou L, Guo Y, Ullah I, Yang Y, Yue Y. 2019. Genome-wide analysis of the callose enzyme families of fertile and sterile flower buds of the Chinese cabbage (Brassica rapa L. ssp. pekinensis). FEBS Open Bio **9 (8):**1432-1449. DOI 10.1002/2211-5463.12685

Repka V, Fischerová, I. & Šilhárová, K. 2004. Methyl jasmonate is a potent elicitor of multiple defense responses in grapevine leaves and cell-suspension cultures. Biologia Plantarum **48:**273–283 DOI https://doi.org/10.1023/B:BIOP.0000033456.27521.e5

Ruan YL, Xu SM, White R, Furbank RT. 2004. Genotypic and developmental evidence for the role of plasmodesmatal regulation in cotton fiber elongation mediated by callose turnover. Plant Physiology **136 (4):**4104-4113. DOI 10.1104/pp.104.051540

Schober MS, Burton RA, Shirley NJ, Jacobs AK, Fincher GB. 2009. Analysis of the (1,3)-beta-D-glucan synthase gene family of barley. Phytochemistry **70 (6):**713-720. DOI 10.1016/j.phytochem.2009.04.002

Shi X, Sun X, Zhang Z, Feng D, Zhang Q, Han L, Wu J, Lu T. 2015. GLUCAN SYNTHASE-LIKE 5 (GSL5) plays an essential role in male fertility by regulating callose metabolism during microsporogenesis in rice. Plant Cell Physiology **56(3):**497-509. DOI 10.1093/pcp/pcu193

Shi ZH, Zhang C, Xu XF, Zhu J, Zhou Q, Ma LJ, Niu J, Yang ZN. 2015. Overexpression of AtTTP affects ARF17 expression and leads to male sterility in Arabidopsis. PLoS One **10 (3):**e0117317. DOI 10.1371/journal.pone.0117317

Song L, Wang R, Zhang L, Wang Y, Yao S. 2016. CRR1 encoding callose synthase functions in ovary expansion by affecting vascular cell patterning in rice. Plant Journal **88 (4):**620-632. DOI 10.1111/tpj.13287

Sun M, Voorrips RE, Steenhuis-Broers G, Van't Westende W, Vosman B. 2018. Reduced phloem uptake of Myzus persicae on an aphid resistant pepper accession. BMC Plant Biology **18 (1):**138. DOI 10.1186/s12870-018-1340-3

Toller A, Brownfield L, Neu C, Twell D, Schulze-Lefert P. 2008. Dual function of Arabidopsis glucan synthase-like genes GSL8 and GSL10 in male gametophyte development and plant growth. Plant Journal **54 (5):**911-923. DOI 10.1111/j.1365-313X.2008.03462.x

Tylewicz S, Petterle A, Marttila S, Miskolczi P, Azeez A, Singh RK, Immanen J, Mähler N, Hvidsten TR, Eklund DM, Bowman JL, Helariutta Y, Bhalerao RP. 2018. Photoperiodic control of seasonal growth is mediated by ABA acting on cell-cell communication. Science **360(6385):**212-215. DOI 10.1126/science.aan8576

Wang X, Lu X, Malik WA, Chen X, Wang J, Wang D, Wang S, Chen C, Guo L, Ye W. 2020. Differentially expressed bZIP transcription factors confer multi-tolerances in Gossypium hirsutum L. International Journal of Biological Macromolecules. **146:**569-578. DOI 10.1016/j.ijbiomac.2020.01.013

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. Genomics, Proteomics & Bioinformatics **8 (1):**77-80. DOI 10.1016/s1672-0229(10)60008-3

Wendel JF, Brubaker C, Alvarez I, Cronn R, Stewart JM. 2009. Evolution and Natural History of the Cotton Genus. Genetics and Genomics of Cotton, 3-22. DOI 10.1007/978-0-387-70810-2_1

Wu SW, Kumar R, Iswanto ABB, Kim JY. 2018. Callose balancing at plasmodesmata. Journal of Experimental Botany **69 (22):**5325-5339. DOI 10.1093/jxb/ery317

Xie B, Wang X, Hong Z. 2010. Precocious pollen germination in Arabidopsis plants with altered callose deposition during microsporogenesis. Planta **231 (4):**809-823. DOI 10.1007/s00425-009-1091-3

Xie B, Wang X, Zhu M, Zhang Z, Hong Z. 2011. CalS7 encodes a callose synthase responsible for callose deposition in the phloem. Plant Journal **65 (1):**1-14. DOI 10.1111/j.1365-313X.2010.04399.x

Yao L, Zhong Y, Wang B, Yan J, Wu T. 2020. BABA application improves soybean resistance to aphid through activation of phenylpropanoid metabolism and callose deposition. Pest Management Science **76 (1):**384-394. DOI 10.1002/ps.5526

Yu CS, Chen YC, Lu CH, Hwang JK. 2006. Prediction of protein subcellular localization. Proteins **64 (3):**643-651. DOI 10.1002/prot.21018

Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D. 2014. CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Research **42 (Database issue):**D1229-1236. DOI 10.1093/nar/gkt1064

Yu Y, Jiao L, Fu S, Yin L, Zhang Y, Lu J. 2016. Callose Synthase Family Genes Involved in the Grapevine Defense Response to Downy Mildew Disease. Phytopathology **106 (1)**:56-64. DOI 10.1094/PHYTO-07-15-0166-R

Zavaliev R, Ueki S, Epel BL, Citovsky V. 2011. Biology of callose (beta-1,3-glucan) turnover at plasmodesmata. Protoplasma **248 (1)**:117-130. DOI 10.1007/s00709-010-0247-0

Hong Z, Delauney AJ, Verma DP. 2001. A cell plate-specific callose synthase and its interaction with phragmoplastin. Plant Cell. **13(4)**:755-68. DOI 10.1105/tpc.13.4.755.

Figure 1

Figure 1. Phylogenetic analysis of CalS protein from *G. hirsutum, G. barbadense, G. raimondii, G. arboretum and Arabidopsis*.

The bootstrap values are shown at the nodes. The *CalSs* from *G. hirsutum*, *G. barbadense*, *G. raimondii*, *G. arboretum*, and *Arabidopsis* are marked with red check, orange rect, purple star, green triangle, grey circle, respectively.

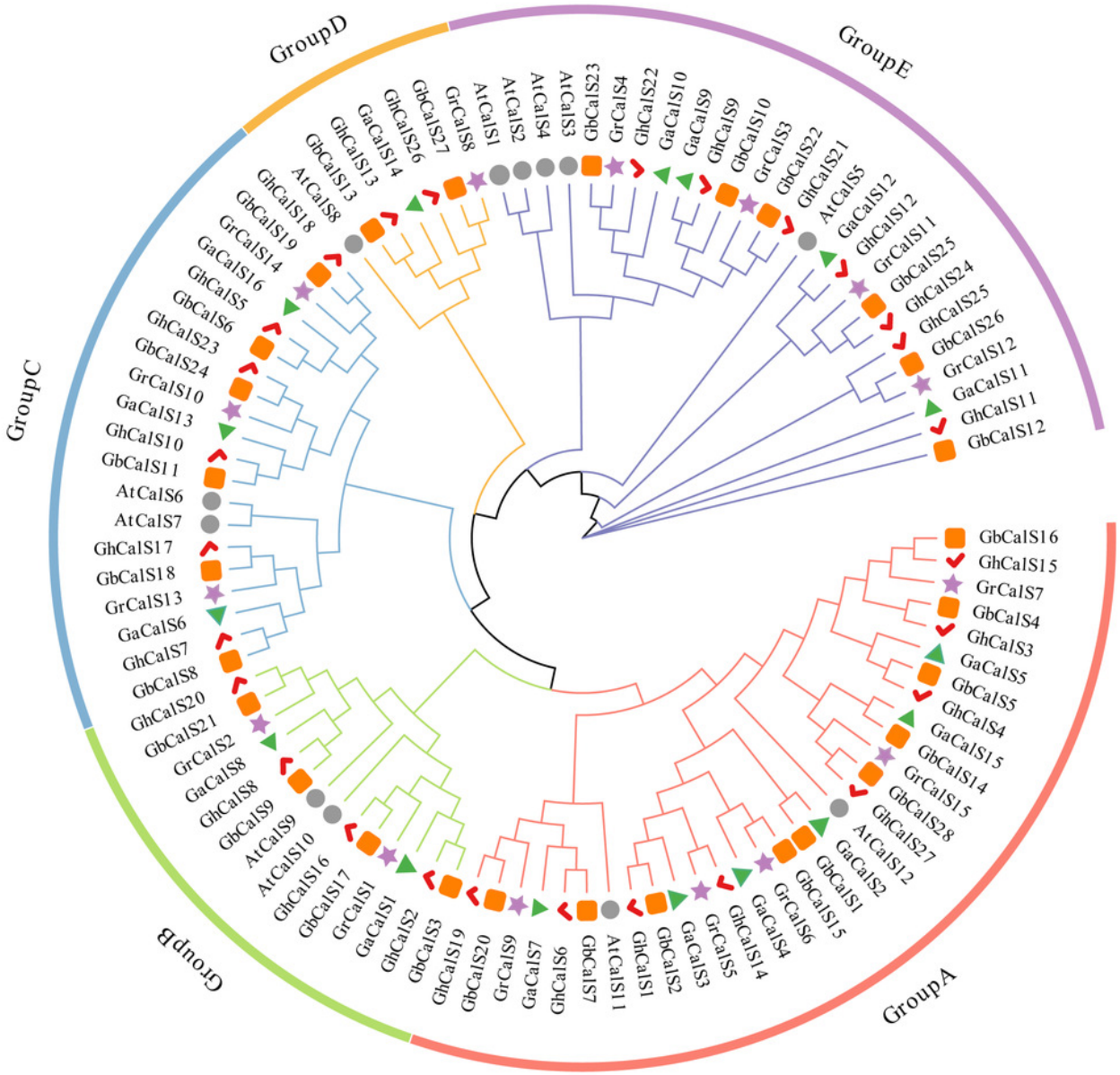
GroupD
GroupE
GroupC
GroupB
GroupA
GbCaIS16
GhCaIS15
GrCaIS7
GbCaIS4
GhCaIS3
GaCaIS5
GbCaIS5
GhCaIS4
GaCaIS15
GbCaIS14
GrCaIS15
GbCaIS28
GhCaIS27
AtCaIS12
GaCaIS2
GbCaIS1
GrCaIS6
GhCaIS14
GaCaIS4
GrCaIS5
GaCaIS3
GbCaIS2
GhCaIS1
AtCaIS11
GbCaIS7
GhCaIS6
GaCaIS7
GrCaIS9
GbCaIS20
GbCaIS19
GhCaIS3
GbCaIS2
GaCaIS1
GrCaIS1
GbCaIS17
GhCaIS16
AtCaIS10
AtCaIS9
GhCaIS9
GaCaIS8
GrCaIS2
GbCaIS21
GhCaIS20
GbCaIS8
GhCaIS7
GaCaIS6
GrCaIS13
GbCaIS18
GhCaIS17
AtCaIS7
AtCaIS6
GbCaIS11
GhCaIS10
GaCaIS13
GrCaIS10
GbCaIS24
GbCaIS23
GhCaIS6
GbCaIS5
GhCaIS16
GaCaIS14
GrCaIS14
GbCaIS19
GhCaIS18
AtCaIS8
GbCaIS13
GhCaIS13
GaCaIS14
GbCaIS26
GbCaIS27
GrCaIS8
AtCaIS1
AtCaIS2
AtCaIS4
AtCaIS3
GbCaIS23
GrCaIS4
GhCaIS22
GaCaIS10
GaCaIS9
GbCaIS9
GbCaIS10
GrCaIS3
GbCaIS22
GhCaIS21
AtCaIS5
GaCaIS12
GbCaIS12
GhCaIS11
GrCaIS11
GbCaIS25
GhCaIS24
GhCaIS25
GbCaIS26
GrCaIS12
GaCaIS11
GhCaIS11
GbCaIS12

Figure 2

Conservative motif and exon-intron structure of CalS genes in cotton.

(A) The evolutionary tree of *CalSs* was constructed using MEGA7. (B) Conservative motif of *CalSs*. The 20 motifs are displayed in different colored boxes. (C) Exon-intron structure of *CalSs*. Introns are presented by grey lines, exons by green boxes, and UTR for yellow boxes.

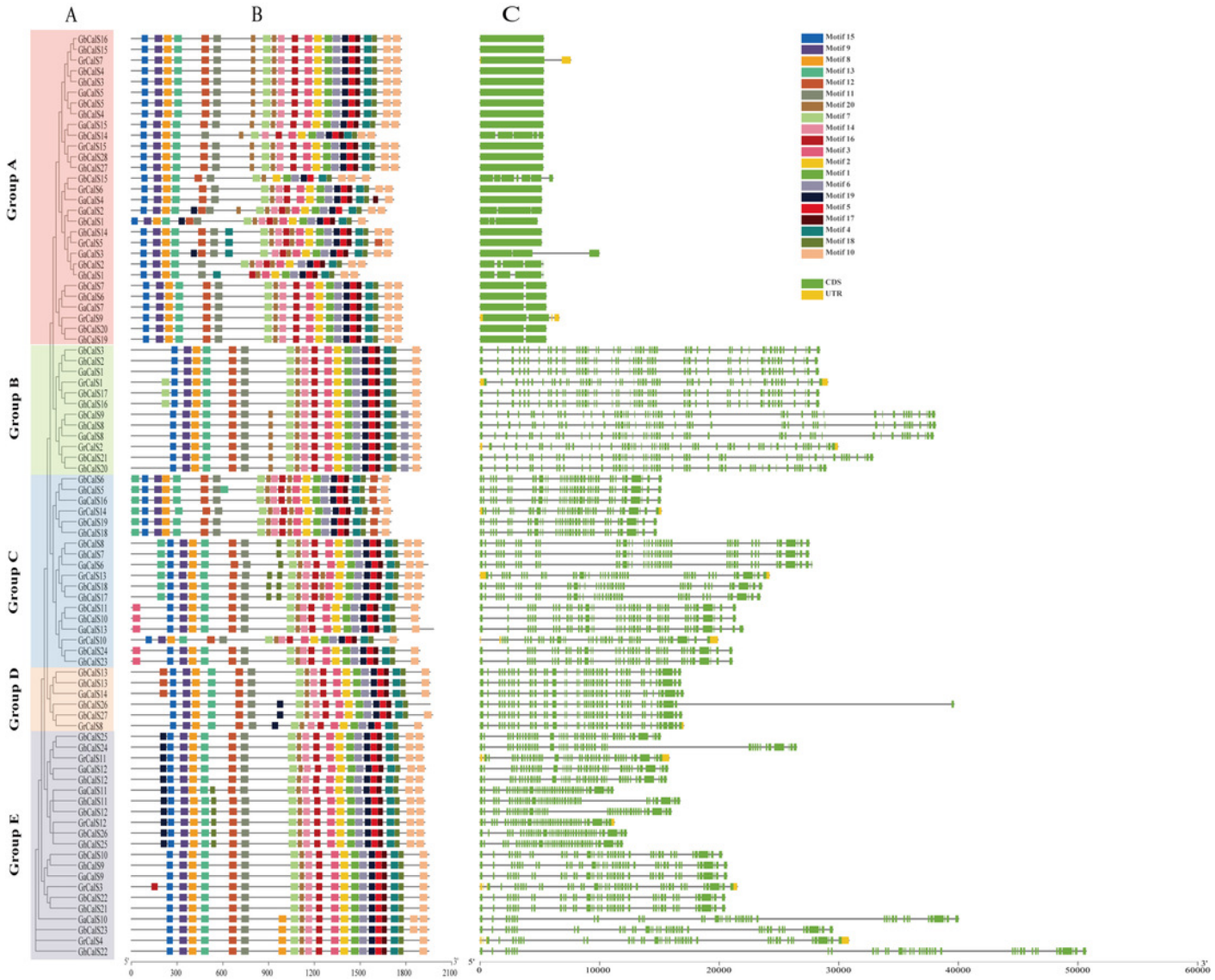

Figure 3

Figure 3. Distribution of 86 *CalSs* on Cotton Chromosomes.

The chromosome name is on the left of each chromosome, and the gene ID is on the right. (A) *G. hirsutum*; (B) *G. barbadense*; (C) *G. arboretum*; (D) *G. raimondii*;

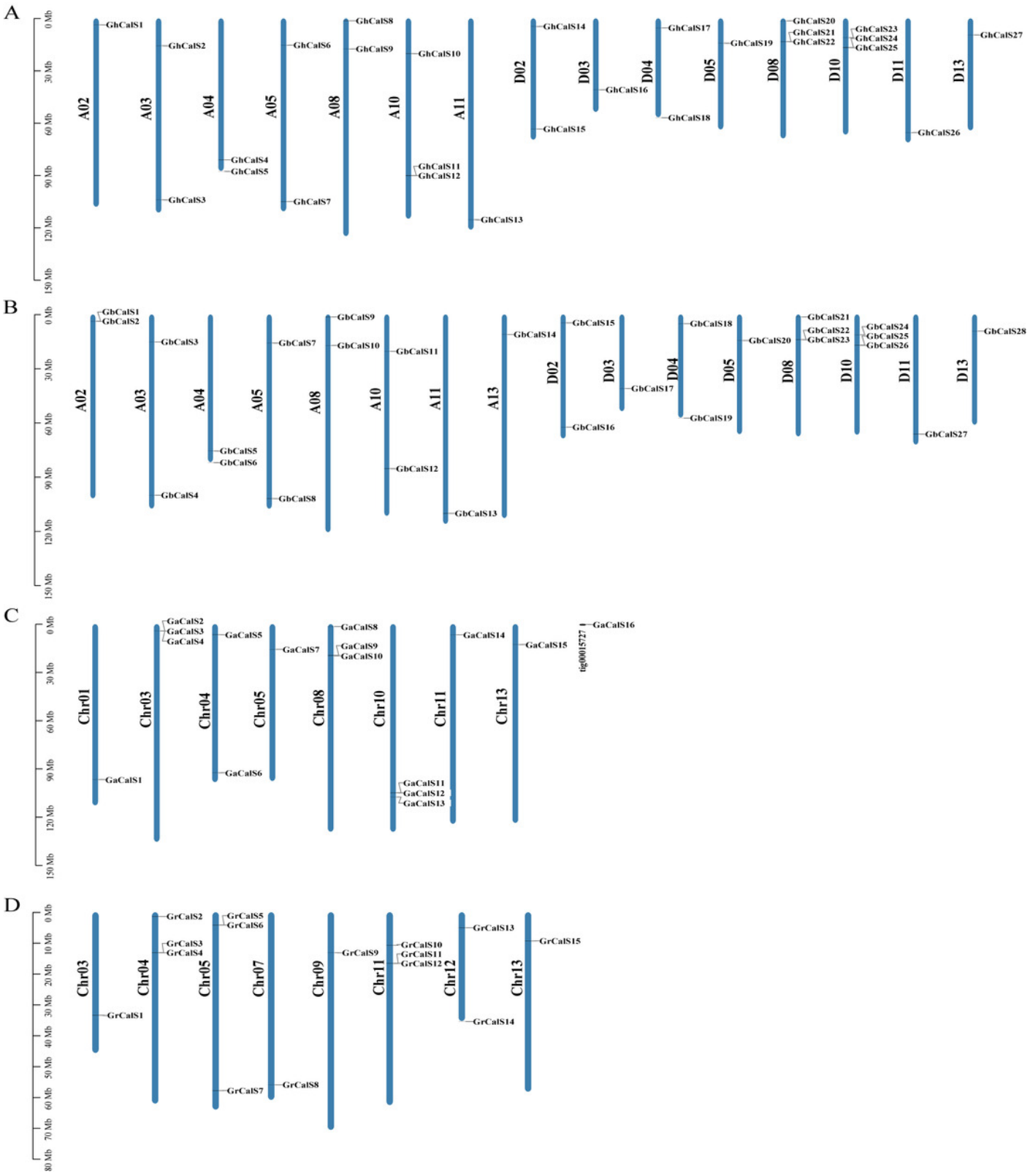
A
A02
A03
A04
A05
A08
A10
A11
D02
D03
D04
D05
D08
D10
D11
D13
GhCalS1
GhCalS2
GhCalS3
GhCalS4
GhCalS5
GhCalS6
GhCalS7
GhCalS8
GhCalS9
GhCalS10
GhCalS11
GhCalS12
GhCalS13
GhCalS14
GhCalS15
GhCalS16
GhCalS17
GhCalS18
GhCalS19
GhCalS20
GhCalS21
GhCalS22
GhCalS23
GhCalS24
GhCalS25
GhCalS26
GhCalS27
0 Mb
30 Mb
60 Mb
90 Mb
120 Mb
150 Mb
B
A02
A03
A04
A05
A08
A10
A11
A13
D02
D03
D04
D05
D08
D10
D11
D13
GbCalS1
GbCalS2
GbCalS3
GbCalS4
GbCalS5
GbCalS6
GbCalS7
GbCalS8
GbCalS9
GbCalS10
GbCalS11
GbCalS12
GbCalS13
GbCalS14
GbCalS15
GbCalS16
GbCalS17
GbCalS18
GbCalS19
GbCalS20
GbCalS21
GbCalS22
GbCalS23
GbCalS24
GbCalS25
GbCalS26
GbCalS27
GbCalS28
0 Mb
30 Mb
60 Mb
90 Mb
120 Mb
150 Mb
C
Chr01
Chr03
Chr04
Chr05
Chr08
Chr10
Chr11
Chr13
GaCalS1
GaCalS2
GaCalS3
GaCalS4
GaCalS5
GaCalS6
GaCalS7
GaCalS8
GaCalS9
GaCalS10
GaCalS11
GaCalS12
GaCalS13
GaCalS14
GaCalS15
GaCalS16
tig00015727 1
0 Mb
30 Mb
60 Mb
90 Mb
120 Mb
150 Mb
D
Chr03
Chr04
Chr05
Chr07
Chr09
Chr11
Chr12
Chr13
GrCalS1
GrCalS2
GrCalS3
GrCalS4
GrCalS5
GrCalS6
GrCalS7
GrCalS8
GrCalS9
GrCalS10
GrCalS11
GrCalS12
GrCalS13
GrCalS14
GrCalS15
0 Mb
10 Mb
20 Mb
30 Mb
40 Mb
50 Mb
60 Mb
70 Mb
80 Mb

Figure 4

Collinearity analysis of *CalSs* in tetraploid and diploid cotton.

(A) Orthologous and paralogous gene pairs among tetraploid and diploid cotton species. The lines represented by various colors indicates the syntenic regions around *CalSs*, and the color between the same species is the same (B) Ka, Ks, Ka/Ks distribution of CalS gene pairs. Ka, Ks, Ka/Ks analysis of *GBCalS-GaCalS*, *GBCalS-GBCalS*, *GBCalS-GhCalS*, *GBCalS-GrCalS*, *GhCalS-GaCalS*, *GhCalS-GhCalS*, *GhCalS-GrCalS*.

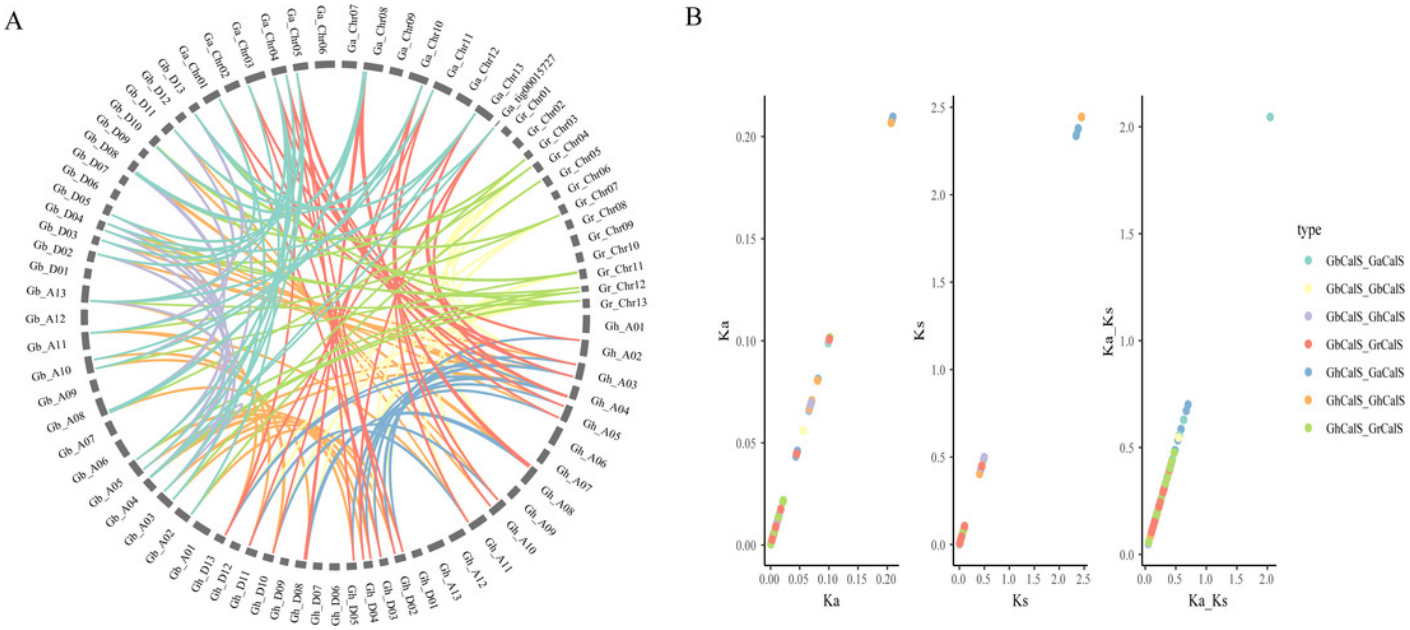

Figure 5

Cis-acting elements on promoters of the *GhCalSs*.

(A)The evolutionary tree of *GhCalSs* was constructed wsing MEGA7. (B) The cis-acting element on the promoter of *GhCalSs*. Number of each cis-acting element in the promoter region.

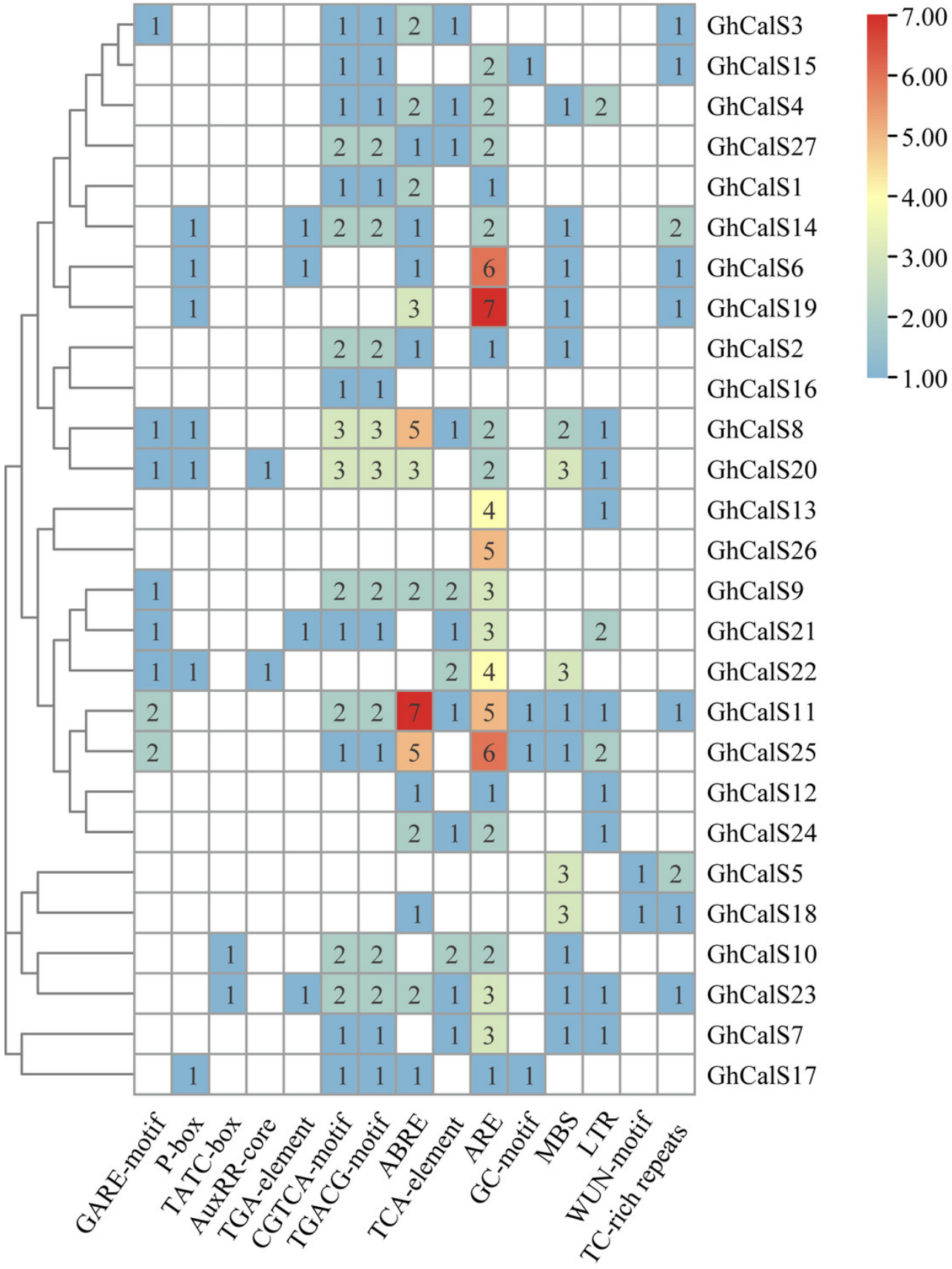
A
B
GhCalS3
GhCalS15
GhCalS4
GhCalS27
GhCalS1
GhCalS14
GhCalS6
GhCalS19
GhCalS2
GhCalS16
GhCalS8
GhCalS20
GhCalS13
GhCalS26
GhCalS9
GhCalS21
GhCalS22
GhCalS11
GhCalS25
GhCalS12
GhCalS24
GhCalS5
GhCalS18
GhCalS10
GhCalS23
GhCalS7
GhCalS17
7.00
6.00
5.00
4.00
3.00
2.00
1.00
GARE-motif
P-box
TATC-box
AuxRR-core
TGA-element
CGTCA-motif
TGACG-motif
ABRE
TCA-element
ARE
GC-motif
MBS
LTR
WUN-motif
TC-rich repeats

Figure 6

RNA sequence profiling of the CalS gene family.

(A) Heatmap displaying expression of expressed *GHCalSs* under hot, cold, NaCl, and PEG treatment (B) Heatmap displaying expression of expressed *GHCalSs* in each tissue.

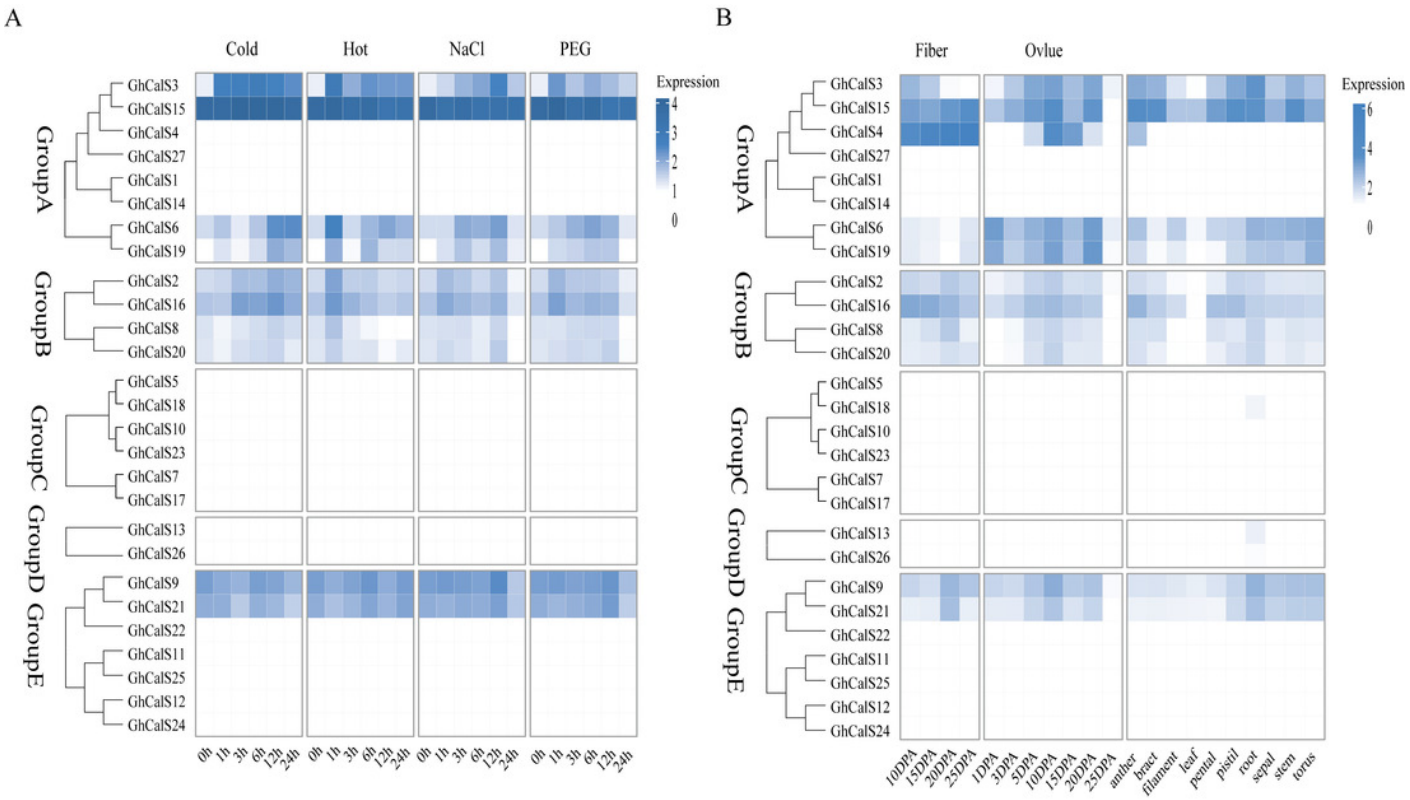

Figure 7

qRT-PCR results of *GHCalSs* under PEG and NaCl.

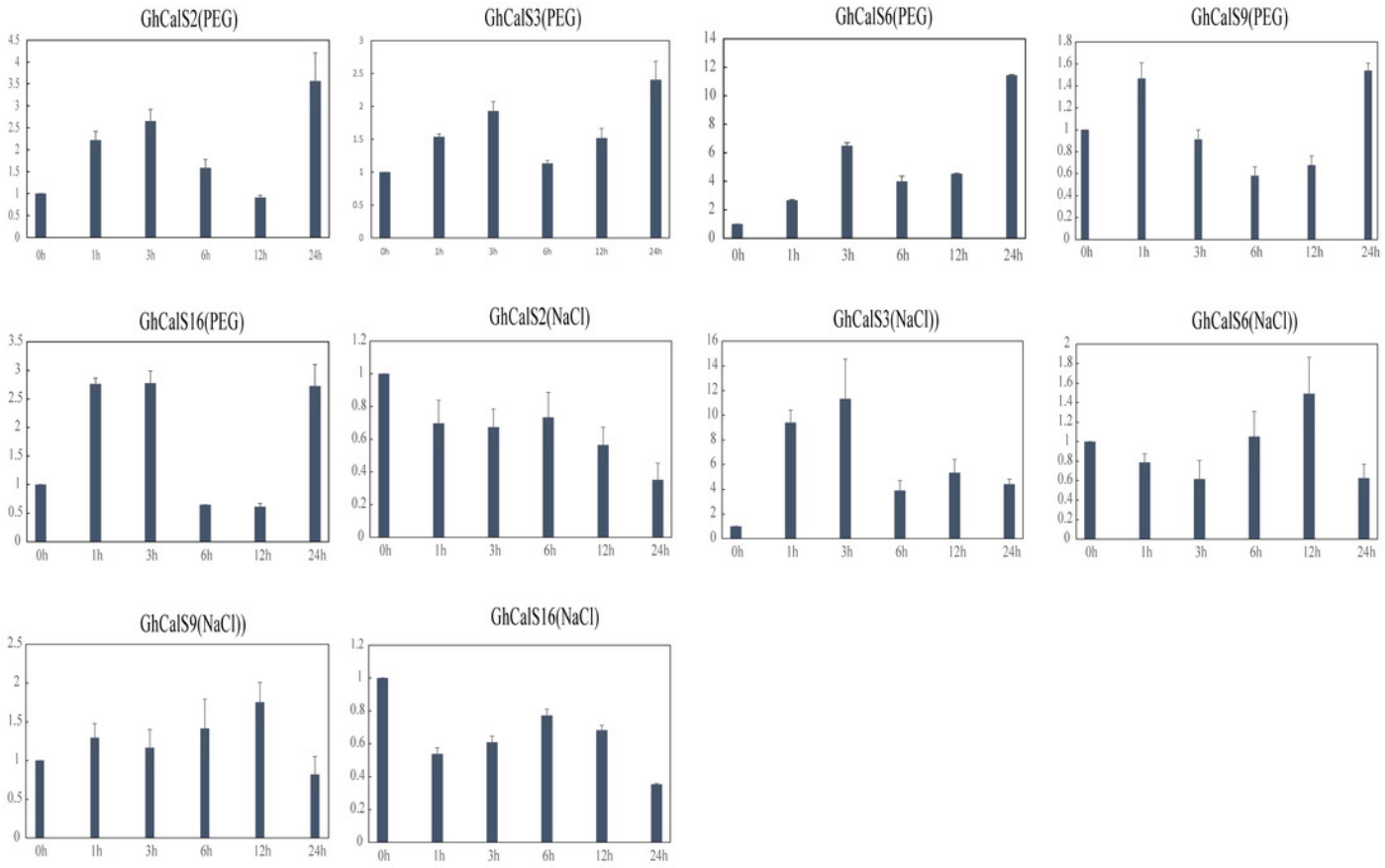

Figure 8

GO enrichment results of *GhCalSs.*

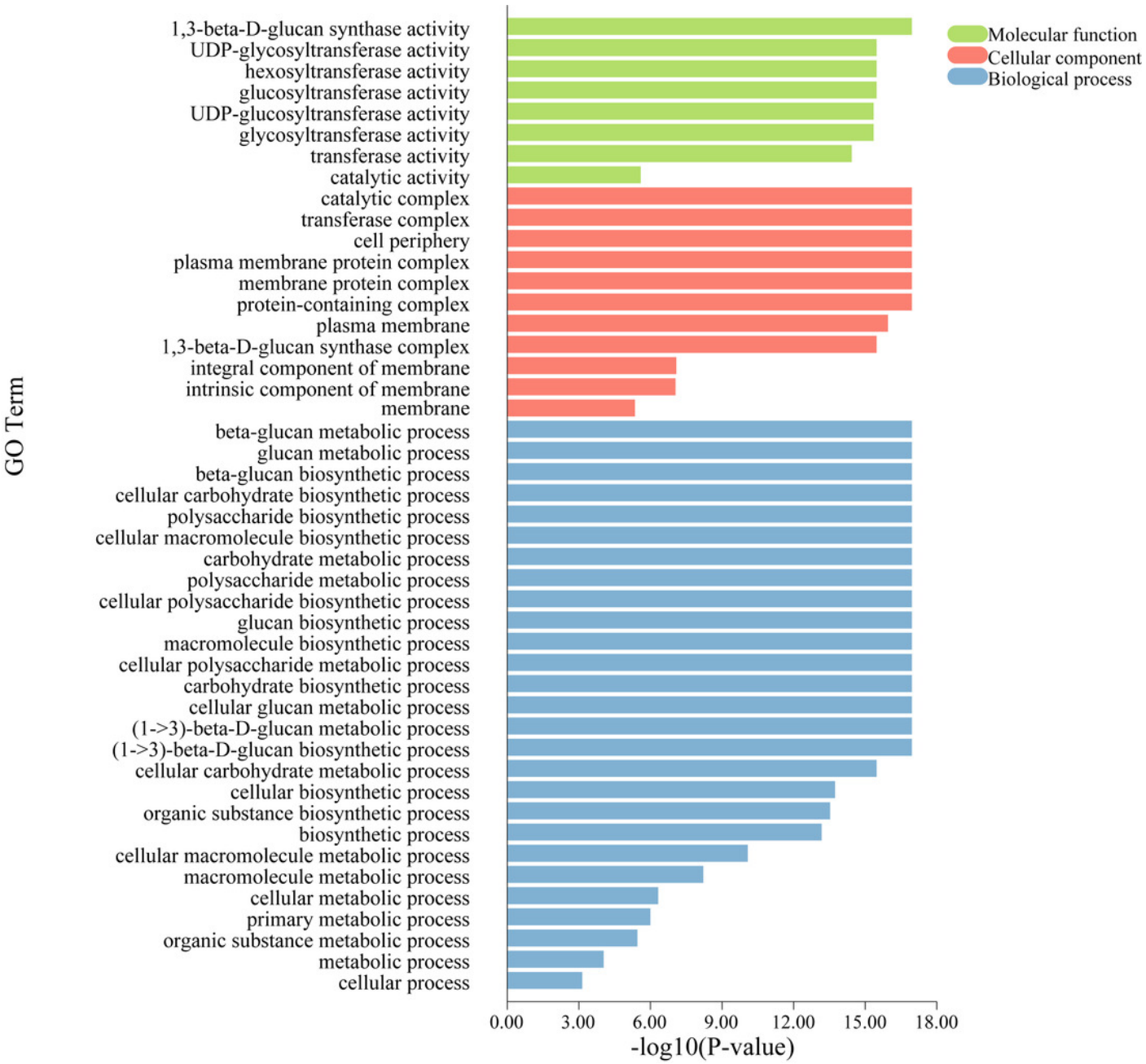

Figure 9

Expression patterns of *GhCalSs* in cotton fiber.

(A)The expression of *GhCalSs* of MBI7747 and CCRI45 at different fiber development stages.

(B) qRT-PCR results of *GhCalS4* at different fiber development stages.