

RESEARCH

Open Access



# Functional structure analysis and genome-wide identification of CNX gene family in cotton

XU Nan, ZHANG Hong, ZHANG Yuexin, FAN Yapeng, WANG Jing, MALIK Waqar Afzal, RUI Cun, HAN Mingge, LU Xuke, CHEN Xiugui, WANG Junjuan, WANG Delong, WANG Shuai, CHEN Chao, GUO Lixue, ZHAO Lanjie and YE Wuwei\*

## Abstract

**Background:** Under abiotic stress conditions, cotton growth is inhibited and yield losses are severe. Identification of calnexin family members and function analysis under abiotic stress laid the foundation for the screening of stress-related candidate genes.

**Results:** A total of 60 CNX family members have been identified in *Gossypium hirsutum*, *G. barbadense*, *G. arboreum*, and *G. raimondii*, and they were divided into two categories: CNX and CRT genes. Through the construction of a phylogenetic tree, they were subdivided into three classes. Further analysis of chromosome localization, conserved promoters, gene structure and selection under pressure showed that the family members were highly conserved in the evolution process. Analysis of *cis*-acting elements in the promoter regions showed that CNX family genes contain regulatory elements for growth and development, anaerobic, drought, defense and stress response, and plant hormones. Using RNA-seq data to study the expression pattern of *GhCNX* genes under cold, hot, salt stress and Polyethylene glycol, it was observed that the gene expression levels changed by different degrees under different stress conditions, indicating that *GhCNX* members were involved in the regulation of multiple biological stresses.

**Conclusion:** This study provides an insight into the members of cotton CNX genes. The results of this study suggested that CNX family members play a role in defense against adversity and provide a foundation for the discovery of stress-related genes.

**Keywords:** Calnexin, *GhCNX*, Conserved motifs, Selection pressure, Collinearity, Differential expression

## Introduction

In eukaryotic cells, the endoplasmic reticulum (ER) is an important organelle for lipid biosynthesis and protein formation. The correct folding and assembly of nascent proteins in the endoplasmic reticulum are monitored by the so-called endoplasmic reticulum quality control,

which only allows correctly folded proteins to leave the endoplasmic reticulum (Nguyen et al. 2019). Once the endoplasmic reticulum homeostasis is disrupted under stresses or by certain environmental signals, and misfolded proteins are accumulated in the endoplasmic reticulum, the quality control of the endoplasmic reticulum will trigger multiple reactions, such as the unfolded protein response (UPR) and endoplasmic reticulum, net-related degradation system (ERAD) (Hwang and Qi 2018; Ruggiano et al. 2014).

\*Correspondence: yew158@163.com

Institute of Cotton Research of Chinese Academy of Agricultural Sciences/ Zhengzhou Research Base, State Key Laboratory of Cotton Biology/School of Agricultural Sciences, Zhengzhou University, Anyang 455000, Henan, China



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

The endoplasmic reticulum is composed of various enzymes and lectin chaperones, and nascent glycoproteins form higher-order structures by means of the folding mechanism of the ER (Schrag et al. 2003). Since the accumulation of structurally incorrect proteins can cause unfolded protein response (UPR) and trigger cellular stresses (Rao and Bredesen 2004), the ER quality control (ERQC) system is very important to reduce the production of misfolded proteins (Sitia and Braakman 2003). The ER chaperones protein calnexin (CNX) and calreticulin (CRT) recognize immature proteins and help them form correct structure (Nakao et al. 2017; Sakono et al. 2014).

CNX family members are ubiquitously found and conserved in all domains of biology from fungi and animal to plants. Plants CNX members belongs to a very unique category of molecular chaperones which interact with newly synthesized membranes and soluble proteins of secretory path. Plants lacking CNX chaperones may break cell homeostasis due to improper protein folding under ER stress and finally lead to plant death. Calnexin plays a vital role in correct folding of proteins and the mechanism of plant resistance, the versatility of CNX may be related to the existence of multiple genes. Knock-out of *CNE1* (sequence similar to calnexin and calreticulin) escalates the secretion of unfolded glycoproteins in *Saccharomyces cerevisiae* signifying their role in processing of nascent proteins (Parlati et al. 1995). There are two types of CNX, CNX1 and CNX2, in *Arabidopsis* which are located in the ER, including the ER in the plasmodesmata, to connect the intercellular channels of plant cells (Liu et al. 2017). Studies have proved that overexpression of CNX in *S. cerevisiae* fission induces cell death using apoptosis markers, and verified the conservation of CNX in apoptosis triggered by ER stress (Guérin et al. 2008). The accumulation of CNX in soybean (*Glycine max* Linn. Merr.) roots was significantly reduced under osmotic or other abiotic stress treatments (Nouri et al. 2012). The absence of CNX and CRT have adverse effects on the root hair length and pollen tube growth of *Arabidopsis* (Vu et al. 2017). Environmental stresses have different effects on the expression of calmodulin genes. The expression of calnexin gene *CNX1* in *S. cerevisiae* fission is induced by hot stress (Jia et al. 2008), while drought stress inhibits the expression level of *NAC2* in soybean hence reduces the ability of the ER to fold the newly synthesized proteins. However, the CNX transcription level in the leaves of the transgenic overexpression plants under drought stress was higher than that of the wild type under the same growth conditions. Overexpression of the CNX gene may reduce the sensitivity of cells to osmotic pressure, enabling plants to survive under severe drought conditions (Valente et al. 2009).

CNX and CRT possibly share a common origin, the calreticulin founder gene within green plants duplicated in early tracheophytes leading to two possible groups of orthologs with specialized functions (Del Bem 2011). Studies have shown that CRT gene exists in the form of multiple copies in plant genome, and CRT gene duplication is a common process in plants (Wasag et al. 2019). In order to study the multifunctionality of CNX genes in plant defense response and development, we conducted a systematic study on CNX gene family to determine the characteristics and phylogenetic relationships among 60 CNX/CRT genes in four *Gossypium* species. Then the phylogenetic analysis, gene structure analysis, chromosome distribution, cis-regulatory element promoter analysis, gene replication, co-expression network analysis, collinearity analysis, selection pressure analysis of repeated gene pairs and subcellular location prediction of GhCNX family members were carried out. In addition, we also determined the functional diversity and expression profile of CNX genes under different abiotic stresses. This may help to clarify the evolutionary mechanism of the cotton CNX gene family, and also help us to further study the stress response genes of cotton and provide valuable information for the breeding of stress-resistant cotton.

The main purpose of this research is to find a new relationship between structure and function, and provide a basis for the functional verification and biological identification of various members of the CNX family, and provide clues for clarifying the specific roles of different types of CNX proteins under different environmental conditions.

## Materials and methods

### Identification of CNX family members

Whole genome files were retrieved from Cotton FGD: *G. arboreum* (CRI), *G. barbadense* (HAU), *G. hirsutum* (NAU) and *G. raimondii* (JGI) (<https://cottonfgd.org/>) (Zhu et al. 2017). The protein sequence of *Arabidopsis* AT5G61790 was found containing calreticulin (PF00262) domain using Pfam database (<http://pfam.xfam.org/>). Hidden Markov Model (HMM) of PF00262 domain was used as the query to screen the candidate genes with calreticulin conserved domains for CNX gene family. Online Pfam database and CD-Search Tool (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) were used for further verification following manual screening of redundant genes. Genes were renamed according to their chromosomal locations and the online tool ExPASy-ProtParam (<https://web.expasy.org/protparam/>) was used to get the physicochemical properties of GhCNXs protein (Gasteiger et al. 2005).

### Phylogenetic analysis and sequence alignment

To investigate the evolutionary relationship among *CNX* members of four *Gossypium* species (*G. arboreum*, *G. barbadense*, *G. hirsutum*, *G. raimondii*), PF00262 (calreticulin) was used as the keyword. The protein sequences of four *Gossypium* species *CNX* family members were entered into MEGAX software with default parameters to construct a maximum likelihood method phylogenetic tree with 1 000 bootstrap replicates. Online web server EvolView (<https://www.evolgenius.info/evolview/>) was also used to improve the visual appearance (Kumar et al. 2016).

### Chromosome localization

The positions of each *CNX* gene on chromosome in four *Gossypium* species were drawn with TBtools software (Chen et al. 2020). Cotton FGD (<http://www.cottonfgd.org/>) (Zhu et al. 2017) was used to retrieve the all required genomic and CDS sequence files.

### Collinearity analysis of *CNX* gene in four *Gossypium* species

Collinearity analysis of *CNX* family members in four *Gossypium* species was performed using TBtools. The evolution of gene families generally undergoes three processes, namely tandem replication, fragment replication and whole genome replication (Xu et al. 2012). The CDS sequence, chromosome length files, genome files and protein files of cotton *CNX* family members were prepared in advance for collinearity analysis. Then, we used the MCScanX module of TBtools software to analyze the collinearity relationship between duplicate gene pairs from four *Gossypium* species (Wang et al. 2012), and used the software for visual analysis (Krzywinski et al. 2009).

### Family members select pressure analysis

TBtools software was used to measure the selection pressure among duplicated gene pairs of four *Gossypium* species using the criteria as given by: the short sequences cover more than 80% of the long sequence after the alignment, and the minimum homology of the aligned region is equal to or greater than 80% (Li et al. 2009). The selection pressure was studied by calculating the nonsynonymous to synonymous ratio ( $Ka/Ks$ ) of duplicate gene pairs.

### Analysis of conserved protein motifs and gene structure

To further understand the relationship between conserved protein motifs and gene structure among cotton *CNX* family members, MEGAX software was used to conduct phylogenetic analysis of *GhCNX* family

members, and then online web server of Motif Enication (MEME) (<https://meme-suite.org/>) to identify conserved protein motifs (Malik et al. 2020).

### Analysis of *GhCNX* promoter region and differentially expressed genes

DNA sequence of 2 000 bp upstream region of *GhCNX* genes was obtained from CottonFGD database and PlantCare website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) was used to get the *cis* regulatory elements present in promotor regions of *GhCNX* genes. The *cis*-acting elements related to plant hormones, plant growth and development, and abiotic stresses were selected for *CNX* family members analysis. To analyze the expression patterns of *GhCNX* gene family members, RNA-seq data (PRJNA490626) downloaded from NCBI database (<https://www.ncbi.nlm.nih.gov/>) were used to analyze the expression level of family genes under cold, hot, salt stress, and Polyethylene glycol (PEG) treatment (Hu et al. 2019).

### Tissue specificity of *CNX* family genes in *G. hirsutum* and analysis of differentially expressed genes under various abiotic stresses

Under controlled environment of 28 °C temperature for 16 h and at 25 °C for 8 h, *G. hirsutum* cv H177 was grown in sandy soil. Stress treatment was carried out at the three-leaf stage of cotton, and sampling was carried out. A total of 1.0~1.5 g of fresh leaves was weighed, quickly transferred to liquid nitrogen for freezing, and stored at -80 °C. Twenty genes of *G. hirsutum* *CNX* family members were tested for tissue specificity (root, stem and leaf). Quantitative Real-time PCR test was completed on the ABIPrism7500 Fast instrument. Three biological repeats of the selected genes were amplified by two-step method. *Actin* (AY305733) was used as the internal reference gene, and  $2^{-\Delta\Delta Ct}$  method was used to calculate the expression value (Zhu et al. 2019). The results were analyzed by GraphPad Prism 8.0 software. Duncan's Multiple Range Test was used to compare the least significant difference of means ( $P < 0.05$ ). The leaf expression data of *GhCNX* family members genes under abiotic stress were obtained from *Gossypium* Resource And Network Database (GRAND) (Zhang et al. 2022). The material was *G. hirsutum* cv TM-1, and the treatment conditions were NaCl (0.4 mol·L<sup>-1</sup>), PEG (200 g·L<sup>-1</sup>), cold, and heat treatment. Then the qRT-PCR of 8 genes under 12% PEG6000 (Beijing Solarbio Science & Technology Co., Ltd.) treatment was represented by heat map.

### Co-expression network analysis of family members

The expression data of four abiotic stresses were used to construct the co-expression network of *CNX* family

members of upland cotton, the correlation coefficient of the data was analyzed, and the data was visualized by Cytoscape software.

### Virus induced gene silencing (VIGS)

Using an upland cotton cultivar H177 as the test material, the *Agrobacterium* solution was injected when two true leaves were unfolded (growing to about 7 days), and the cotton was treated with 12% PEG6000 to simulate drought stress during the three-leaf stage. VIGS needs appropriate positive control and negative control to observe the effect of silence. The positive control usually adopts phytoene desaturase (*PDS*) gene silencing. *PDS* is a key enzyme in the carotenoid synthesis pathway. When the expression of *PDS* is blocked, the plant loses the photoprotective effect of carotenoids, thus showing an albino effect. Negative control generally adopts virus free empty body (pYL156). The qRT-PCR method mentioned in section “Tissue specificity of *CNX* family genes in *G. hirsutum* and analysis of differentially expressed genes under various abiotic stresses” is used to evaluate the effect of silence.

## Result

### Identification of *CNX* gene family

A total of 60 family members have been identified for the four *Gossypium* species including 20 in *G. hirsutum* (*GhCRT1-12* and *GhCNX1-8*), 20 in *G. barbadense* (*GbCRT1-12*, *GbCNX1-8*), 10 in *G. arboreum* (*GaCRT1-6*, *GaCNX1-4*), 10 in *G. raimondii* (*GrCRT1-6*, *GrCNX1-4*). Genes were renamed according to chromosome position information (Additional file 1: Table S1). The number of *CNX/CRT* in tetraploid *Gossypium* species (*G. barbadense* and *G. hirsutum*) is double as compared with diploid ones (*G. arboreum* and *G. raimondii*).

The physicochemical properties of *CNX* proteins in *G. hirsutum* were analyzed, and the results showed that the isoelectric point (pI) ranges from 4.44 (*GhCRT11*) to 6.27 (*GhCRT10*). The minimum number of amino acids was found 417 (*GhCRT9*), and the maximum 541 amino acids (*GhCNX7*), and the molecular mass ranges from 47.87 kDa (*GhCRT9*) to 61.59 kDa (*GhCNX7*) (Additional file 2: Table S2).

In order to study the evolutionary relationship of *CNX* genes in plants, the protein in sequences of the above four *Gossypium* species were aligned and phylogenetic trees were constructed. In order to further investigate the *CNX* genes from four *Gossypium* species, MEGAX was used to construct an unrooted phylogenetic tree between the four *Gossypium* species through the neighbor-joining method (Fig. 1). The family is divided into three classes: a, b, and c. Clade a mainly contains the *CNX* genes, clade b contains the *CRT1*, *CRT2*, *CRT7*, and *CRT8* genes of

four *Gossypium* species, clade c contains the remaining *CRT* genes. In the *CNX* family, the genetic relationship between *G. hirsutum* and *G. barbadense* is closer than that between *G. arboreum* and *G. raimondii*.

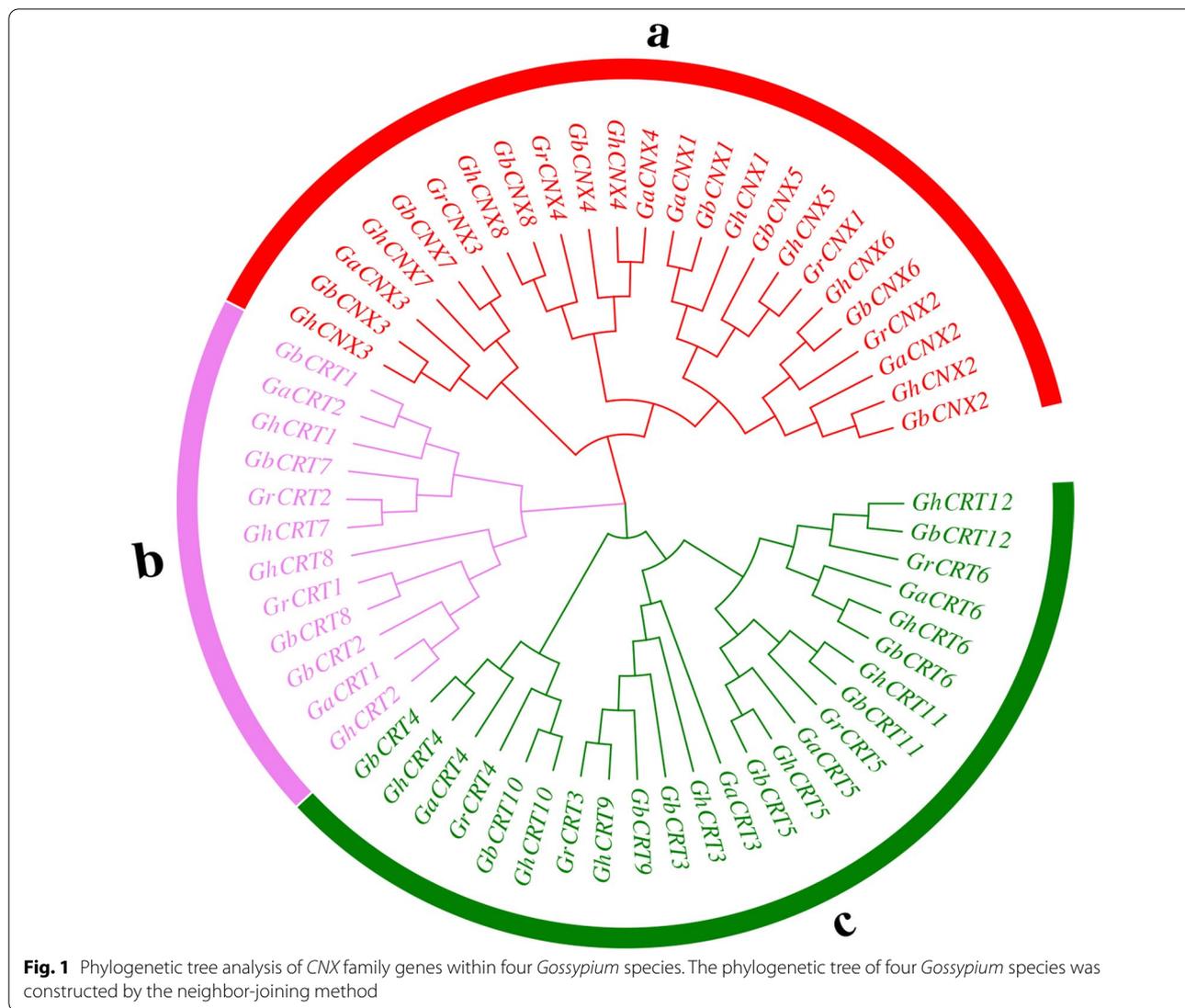
### Chromosome position analysis of *CNX* family

In order to further analyze the characteristics of *CNX* family member genes, chromosome location analysis was carried out for the members of this gene family in four *Gossypium* species (Fig. 2). *CNX* members of *G. hirsutum* and *G. barbadense* showed the similar trend of localization. Analysis of the chromosomal distribution of *G. arboreum* in this gene family revealed that the A genome was identical to the At-subgenome of *G. hirsutum* and *G. barbadense*, except for chromosomes 2 and 3. It can be seen that the *GaCRT2* gene on chromosome 3 has been relatively newly added, while a gene on chromosome 2 has been lost. The distribution of *CNX* genes on chromosomes of *G. raimondii* showed obvious gene addition and loss. In the chromosome gene distribution of *G. raimondii*, except chromosome 3 and 13, the distribution of *CNX* genes on the other chromosomes of Dt-subgenome of two tetraploid species was different.

### Collinearity analysis of *CNX* family genes

To further understand the evolutionary relationship of *CNX* genes in cotton, the collinearity analysis of *CNX* family members in four *Gossypium* species was performed using TBtools software. The *CNX* genes of *G. hirsutum*, *G. barbadense*, *G. raimondii* and *G. arboreum* were carried out, and the gene duplication and collinearity between them were analyzed. The *CNX* genes of *G. hirsutum* (*Gh*) and *G. barbadense* (*Gb*) are duplicated between *G. arboreum* (*Ga*) and *G. raimondii* (*Gr*), indicating that the two tetraploid genomes (*G. hirsutum* and *G. barbadense*) arise from two diploids genomes during evolution. Homologous sequences can usually be predicted by genomic collinearity analysis, and some functions of homologous sequences may be similar.

The genes interconnected with collinearity lines represent the duplication relationship between them. Figure 3 shows that many duplicated genes from *GhAt/GhDt*, *GbAt/GbDt* subgenome and *GaA* and *GrD* genomes are connected by lines of the same color, namely the *GhAt/GhDt* and *GbAt/GbDt* subgenomes have *CNX* homologous genes in the *GaA* and *GrD* genomes, indicating that these genomes/subgenomes have close relationship during evolution, and most of the *CNX* genes remained conserved during polyploidization except of few members. Genes located on the same chromosomal region (e value < 1e−5) are classified as tandem duplications, while the rest of the genes from the same genome are considered to be segmental duplications. Overall, 5 tandem, 67



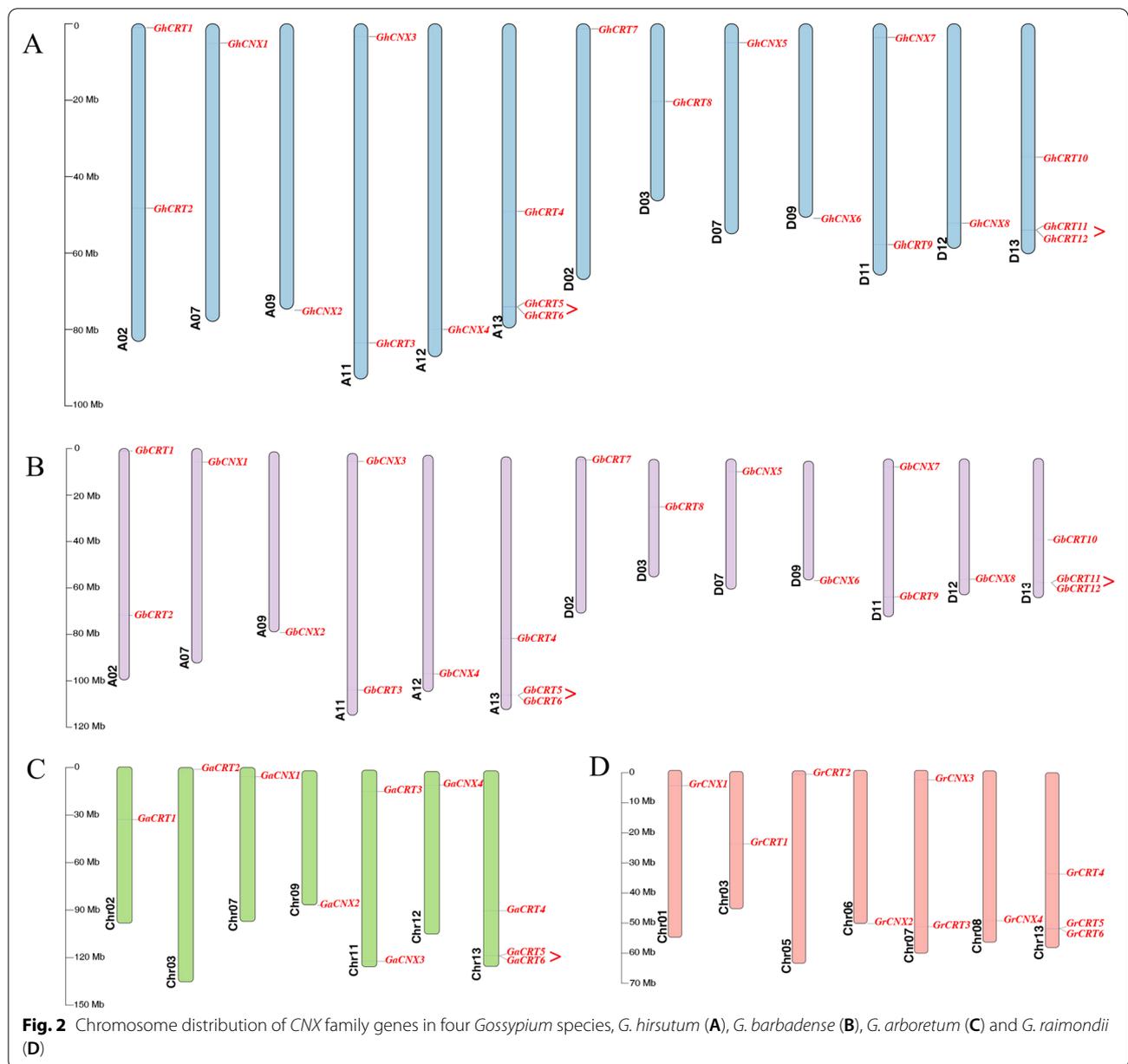
**Fig. 1** Phylogenetic tree analysis of CNX family genes within four *Gossypium* species. The phylogenetic tree of four *Gossypium* species was constructed by the neighbor-joining method

segmental, and 180 whole genome duplicated pairs were identified from four *Gossypium* species. The five tandem duplications comprise of two tandem duplications pairs in *Gh-Gh* and *Gb-Gb* self-alignment, and one tandem duplications in *Ga-Ga* self-alignment. All tandem duplications are located on chromosome 13. It further verifies that the origin of tetraploid *Gossypium* species from diploid ones (Malik et al. 2020). In the self-alignment of the four *Gossypium* species, a total of 67 segmental duplications, and 180 whole-genome duplicated pairs were obtained from collinearity analysis (Fig. 3).

In the process of evolution, the duplicated gene pair may also involve deletions from its original function, which eventually leads to non-functionalization (loss of original function), sub-functionalization (division of original function), and new functionalization (acquisition of new function) (Prince and Pickett 2002). To check the

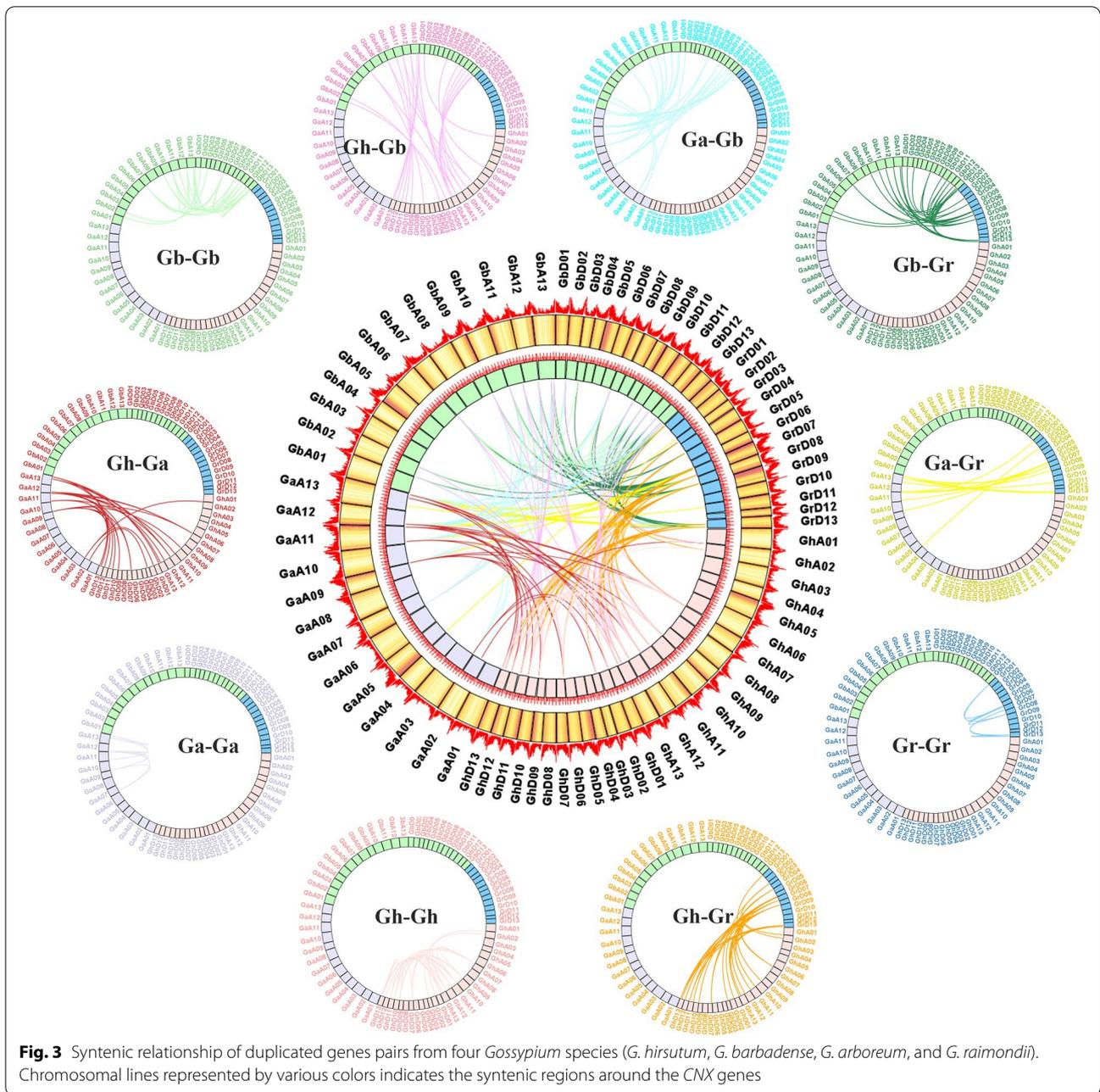
role of selection pressure analysis in evolution, CNX family gene pairs of four *Gossypium* species were analyzed to determine whether there was selection pressure role of CNX family genes. In order to determine the nature and degree of selection pressure on these duplication gene pairs, the values of non-synonymy ( $Ka$ ) and synonymous replacement ( $Ks$ ) of 250 duplication gene pairs in four *Gossypium* species were calculated. These combinations include *Gh-Gh*, *Gh-Gb*, *Gh-Ga*, *Gh-Gr*, *Gb-Gb*, *Gb-Ga*, *Gb-Gr*, *Ga-Ga*, *Ga-Gr*, *Gr-Gr*. It is generally believed that  $Ka/Ks=1$  means neutral selection (pseudogene),  $Ka/Ks < 1$  means purification or negative selection (purification selection effect), while  $Ka/Ks > 1$  means positive selection effect (Wang et al. 2009).

From a total of 250 CNX gene pairs in four *Gossypium* species, only 3 pairs underwent positive selection while other 247 pairs experienced purifying selection pressure



during evolution illustrating their great evolutionary conservation. Gene pairs with positive selection pressure belongs to *Ga–Gb*, and *Ga–Gh* with 1 and 2 pairs each, indicating the relax selection criteria for evolution of *G. hirsutum* as compared with *G. barbadense* but overall selection criteria for both tetraploid species was very tight. The *Ka/Ks* ratio in the number of *Ga–Gr*, *Gb–Gh*, *Gh–Gr*, *Gh–Gh* duplicated gene pairs in the range from 0.5 to 0.99 was 1, 3, 3, and 1, respectively, and the number of duplicated gene pairs with *Ka/Ks* ratios and from 0 to 0.49 was 14, 19, 30, and 7, respectively (Table 1). Overall, 250 pairs of duplicate gene pairs from four *Gossypium*

species (*Gh*, *Gb*, *Ga*, and *Gr*) were observed based on selection pressure analysis. Among them, 247 pairs (98.8%) had *Ka/Ks* values less than 1, including 230 pairs of genes with *Ka/Ks* values less than 0.5 and 17 pairs of genes between 0.5 and 0.99, showing intense purifying selection. Only 3 pairs (1.2%) of repetitive gene pairs have *Ka/Ks* values greater than 1. These gene pairs may have undergone rapid evolution after repetition and have experienced positive selection pressure. Since most of the *Ka/Ks* values are less than 1.0, speculated that the cotton CNX gene family has experienced strong purification selection pressure and limited functional differentiation



after segmental and whole genome duplication (Hurst 2002) (Fig. 4).

**CNX family conserved domains and gene structure analysis**

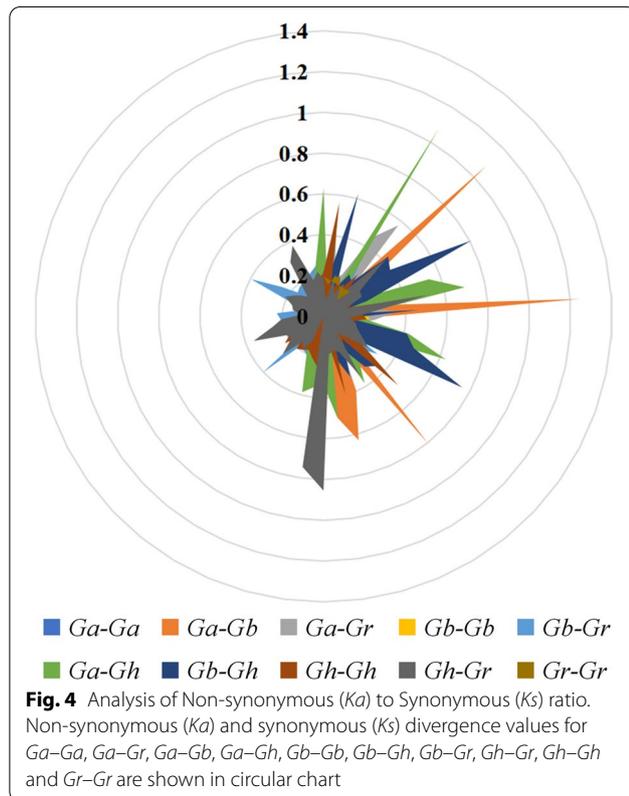
The conserved protein motif domain file obtained from MEME website was combined with the GFF files of the four *Gossypium* species in the TBtools software to obtain the structure and classification specific information of the four *Gossypium* species CNX family (Fig. 5). There are 10 motifs in the CNX domains of the four

*Gossypium* species. According to the combination of phylogenetic tree and motif, the CNX family of the four *Gossypium* species were divided into three categories: I, II, and III. Class I and Class II are mainly composed of CRT genes. CRT was divided into two classes according to conserved motifs and the evolutionary tree. Class III mainly includes CNX genes. The overall structure of CRT genes is consistent. Class I contains 8 motifs. There are 7 motifs in Class II and one motif is less than that in Class I. It is speculated that a certain function has been lost in

**Table 1** Distribution of duplicate gene pairs in selection pressure analysis from four *Gossypium* species

Pairs	Positive selection no	Purify selection NO		Total	Purity/%
		0.99–0.50	0.49–0.00		
Ga-Ga	0	0	9	9	100
Ga-Gr	0	1	14	15	100
Ga-Gb	1	4	21	26	96.15
Ga-Gh	2	5	20	27	92.59
Gb-Gb	0	0	25	25	100
Gb-Gh	0	3	19	22	100
Gb-Gr	0	0	42	42	100
Gh-Gr	0	3	43	46	100
Gh-Gh	0	1	30	31	100
Gr-Gr	0	0	7	7	100
Total	3	17	230	250	98.80

From the point of view of gene structure introns and exons, all genes contain exons and introns. At the same time, the gene structures of Class I, Class II and Class III have their own consistent characteristics. The first part of the *CRT* gene has more exons and introns, and the length of the exons are shorter; the second part is the structure of *CRT* gene, which has a large number of exons and a short length, and the length of *GhCRT8* gene is the longest, mainly because there is a very long intron; *CNX* genes are mainly distributed in the third part, with fewer exons than *CRT* genes, *CRT* contains a total of 14 exons, whereas *CNX* has only 6, and the distribution of exons and introns of *CNX* genes is more uniform, and the whole is more consistent. In short, there are obvious structural differences and unique characteristics among the various components of the *CNX* gene family, and they are relatively conservative.



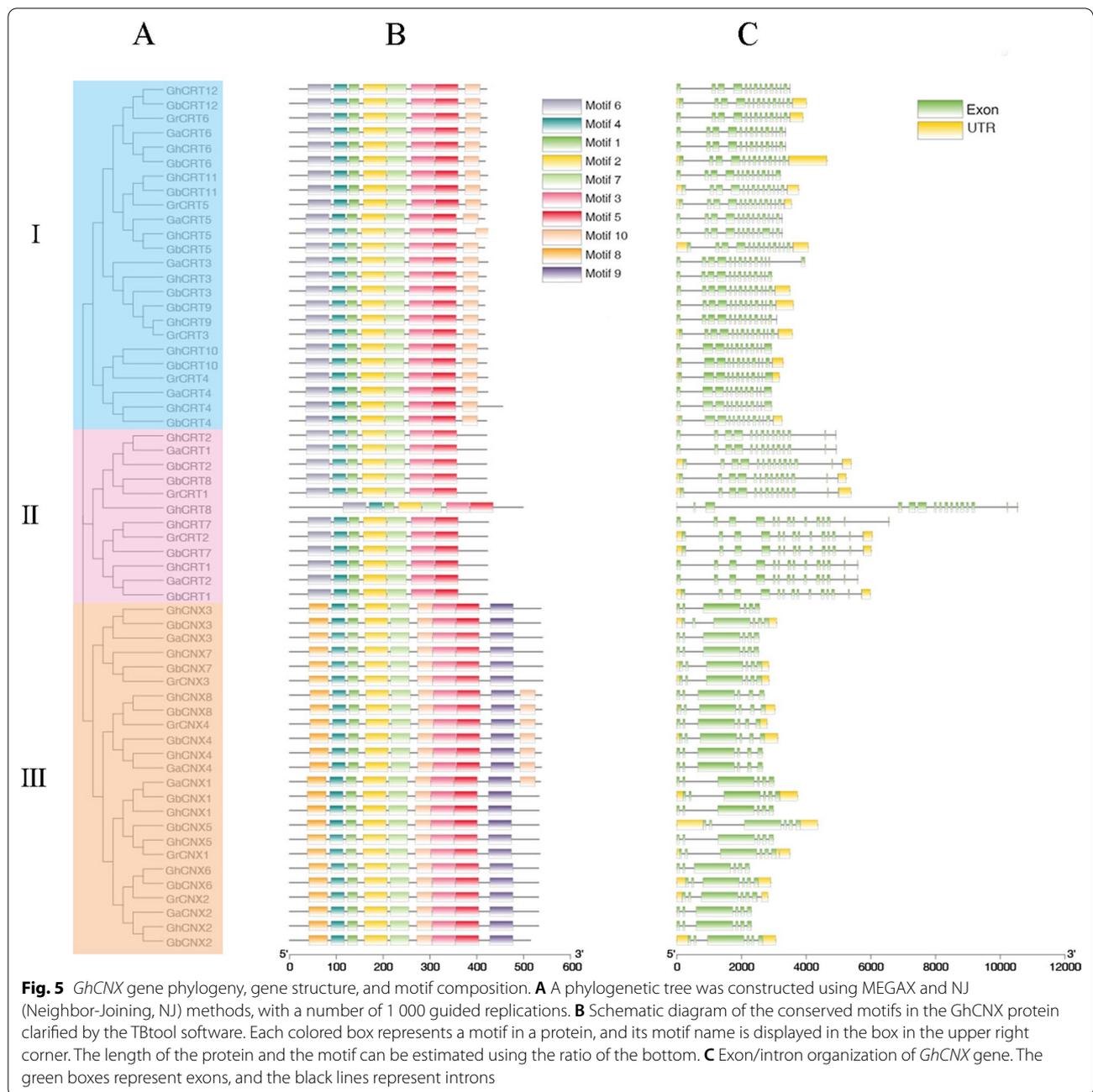
**Analysis of promoters and differentially expressed genes in the *CNX* family of *G. hirsutum***

The members of the *CNX* gene family play an important role in various important physiological and biochemical processes of plants (Fan et al. 2009). In addition, *CNX* is also involved in the response to various environmental stimuli (Oelze et al. 2014). In order to determine the function of the *GhCNX* gene family in different environments, the expression profiles of the *GhCNX* genes during cotton growth and development stages and its response to phytohormones and abiotic stresses were studied (Fig. 6).

*Cis*-acting elements are located in the promoter region of genes and serve as references for tissue-specific and stress responses under different stresses. The *cis*-acting elements of the *CNX* gene family mainly include *cis*-acting regulatory elements involved in Methyl Jasmonate (MeJA) reaction; *cis*-acting elements involved in abscisic acid, gibberellin and auxin plant hormones; anaerobic, drought, low temperature defense and stress response *cis*-acting regulatory elements. The number of *cis*-acting elements of each gene varies, for example, *GhCRT3* contains *cis*-acting element involved in anaerobic-induced, meristem expression, MeJA reaction, abscisic acid and growth related to plant hormones, and defense and stress response; *GhCNX2* contains *cis*-acting elements involved in MeJA reaction, abscisic acid, anaerobic-induced elements, gibberellin reaction and meristem expression. In general, the *CNX* family members of *G. hirsutum* mainly contains *cis*-acting elements related to plant hormones, growth and development, and adversity.

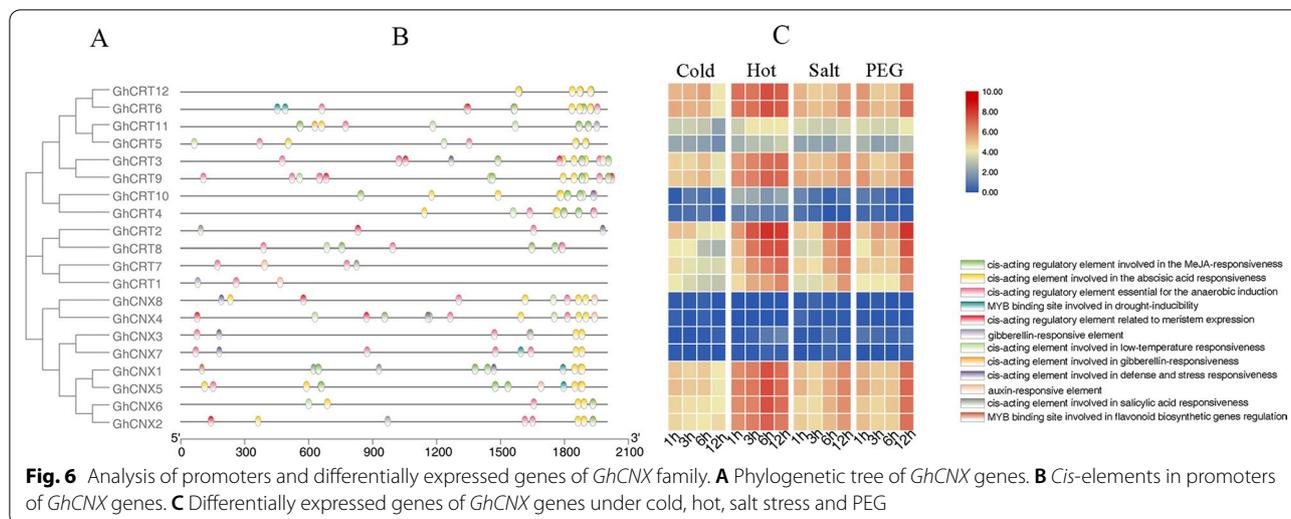
Gene expression patterns provide the important references for gene function analysis, and gene expression is related to the biological functions controlled by *cis*-acting elements. Based on the published RNA-seq data of

the evolutionary process. There are consistent structural domains in each category, with obvious structural characteristics. The third category comprises of *CNX* genes. The *CNX* genes in the classification had the same motif structure. Seven *CNX* family members in the category have one more motif3 than the other.



cotton under different adversity the *GhCNX* gene expression profile under cold, hot, salt, and PEG treatment conditions at 1 h, 3 h, 6 h, and 12 h were analyzed (Wang et al. 2020). As can be seen in the heatmap, there were four parts of the *CNX* family genes (*GhCRT12–GhCRT6*, *GhCRT3–GhCRT9*, *GhCRT2–GhCRT1*, *GhCNX1–GhCNX2*) showed relatively significant differential expression under different stress treatments. For example, the four genes contained in *GhCNX1* to *GhCNX2* have relatively consistent expression patterns under

stress. Under hot stress, the expression level is the highest at 6 h and 12 h. However, the expression levels under the salt and PEG conditions showed a trend from low to high, and the highest expression level is at 12 h. It can be seen that the expression of genes with close evolutionary relationship tends to be consistent under different stresses. For example, four genes, *GhCNX1*, *GhCNX5*, *GhCNX6* and *GhCNX2*, had the highest expression levels at 6 h under heat stress, while their expression levels tended to increase with the treatment time under salt



stress and PEG. These results suggest that their expression may be regulated by abiotic stresses. The *cis*-acting elements predicted by combining the promoter region include MYB (involved in drought-inducibility), anaerobic, abscisic acid, gibberellin-responsiveness and *cis*-acting element involved in defense and stress response. The presence of these *cis*-acting elements in gene promoters suggests that they may participate in the mechanism of stress and may improve abiotic stress response.

**Analysis of tissue specificity and abiotic stresses expression of CNX family members in *G. hirsutum***

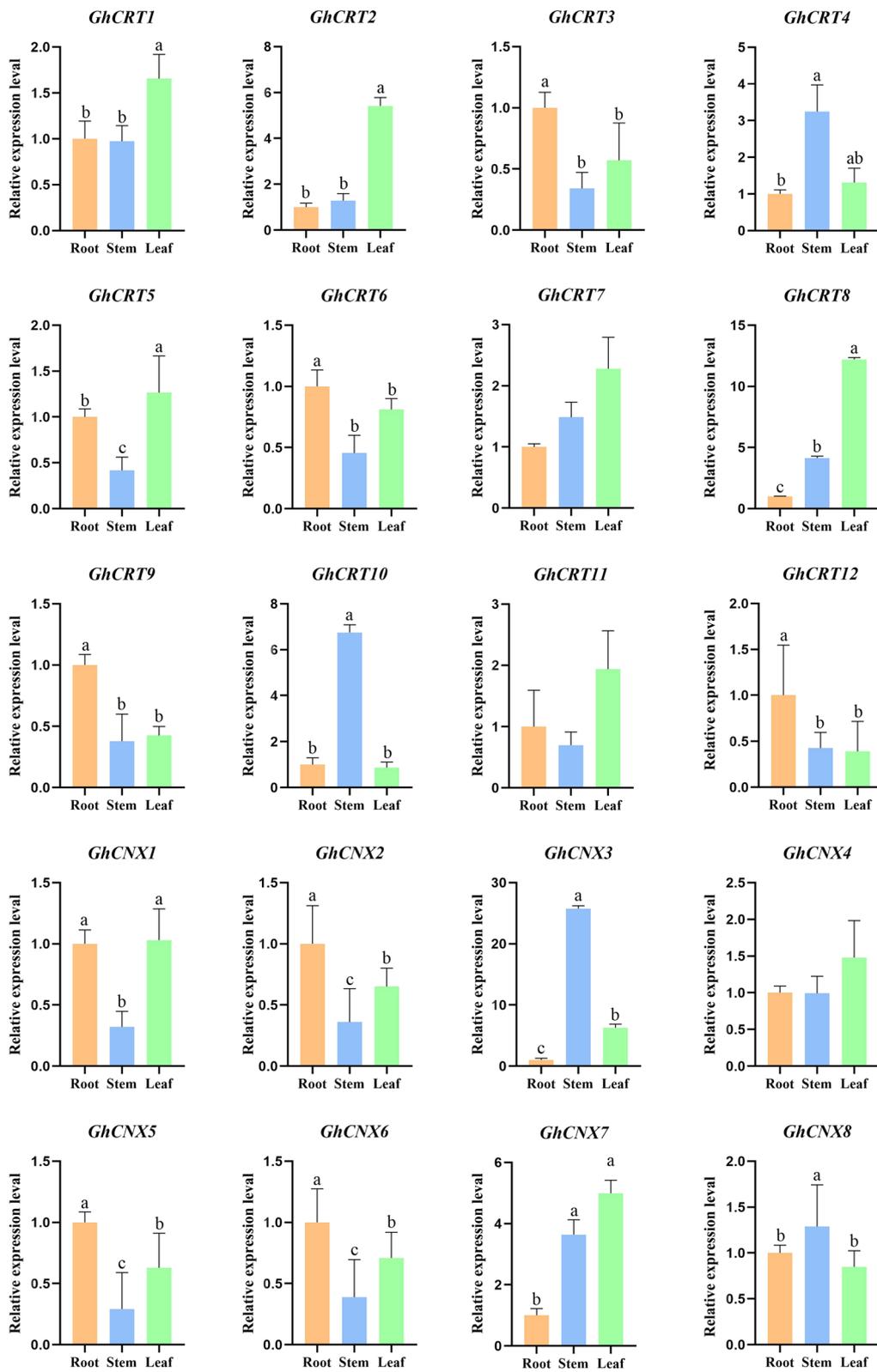
The tissue specificity of 20 *GhCNX* genes were analyzed by expression data, we can see the differential expression of CNX family members in different tissues (Fig. 7). The four genes *GhCRT3*, *GhCRT6*, *GhCRT9*, and *GhCRT12* have the highest expression levels in roots; *GhCRT4*, *GhCRT10*, *GhCNX3*, *GhCNX7*, and *GhCNX8* have the highest expression levels in stems; *GhCRT1*, *GhCRT2*, and *GhCRT8* have the highest relative expression levels in leaves and the 5 genes *GhCNX1*, *GhCRT5*, *GhCNX2*, *GhCNX5*, and *GhCNX6*, have relatively high expression levels in roots and leaves; the expression levels of *GhCRT7*, *GhCRT11*, and *GhCNX4* are not significantly different among the tissues.

Eight genes with high expressions in leaves relative to roots and stems were selected to analyze the expression under 4°C, 37°C, PEG, and NaCl stress, respectively. Eight *GhCNX* family members responded to abiotic stress to varying degrees were found (Fig. 8). Under 4°C stress, *GhCNX2*, *GhCNX5*, and *GhCNX6* genes were down regulated. *GhCNX6*, *GhCRT1*, and *GhCRT5* genes were up-regulated under 37°C. Under PEG treatment, *GhCNX2* was significantly down-regulated, while

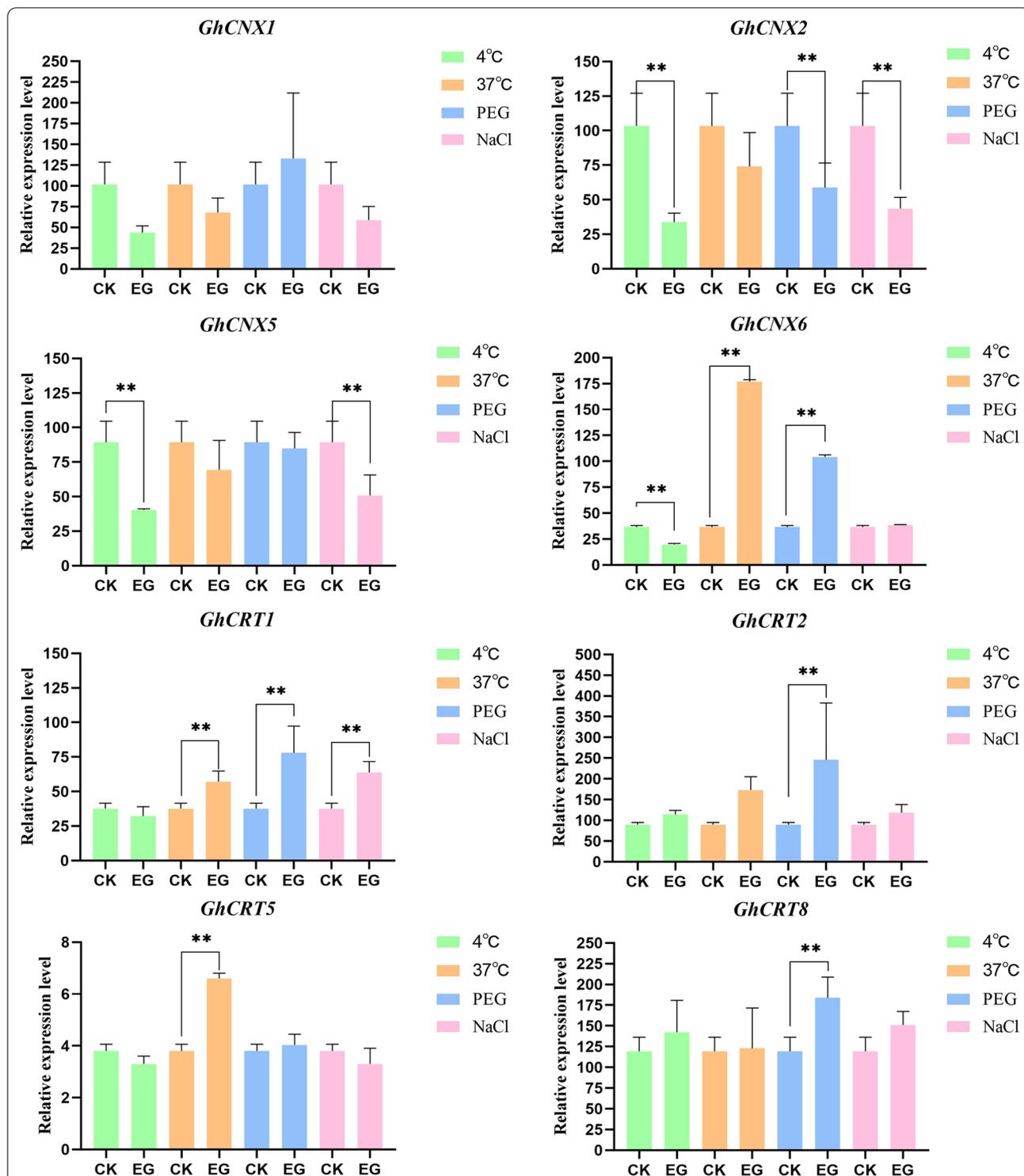
*GhCNX6*, *GhCRT1*, *GhCRT2*, and *GhCRT8* were significantly up-regulated. Under NaCl stress, the expression levels of *GhCNX2*, *GhCNX5*, and *GhCRT1* changed significantly. Among them, *GhCRT1* showed highly up-regulated expression, while *GhCNX2* and *GhCNX5* showed down-regulated expression. Then, 8 genes with relatively high expression levels in leaves were selected for Quantitative Real-time PCR test under 12% PEG6000 treatment (Fig. 9). After treated by 12% PEG6000, the overall expression level increased, and the highest expression level was at 6 h after treatment, then decreased after 12 h treatment. It can be seen that the differential expression of different genes has a certain relationship with the genetic relationship. The expression levels among *GhCNX1*, *GhCNX2*, *GhCNX5*, and *GhCNX6* were the same. After 6 h of 12% PEG6000 treatment, the expression levels of *GhCNX1* and *GhCNX6* were higher than other genes, and the expression of CNX gene was different in stress treatment.

**Co-expression network of GhCNXs family members under four abiotic stresses**

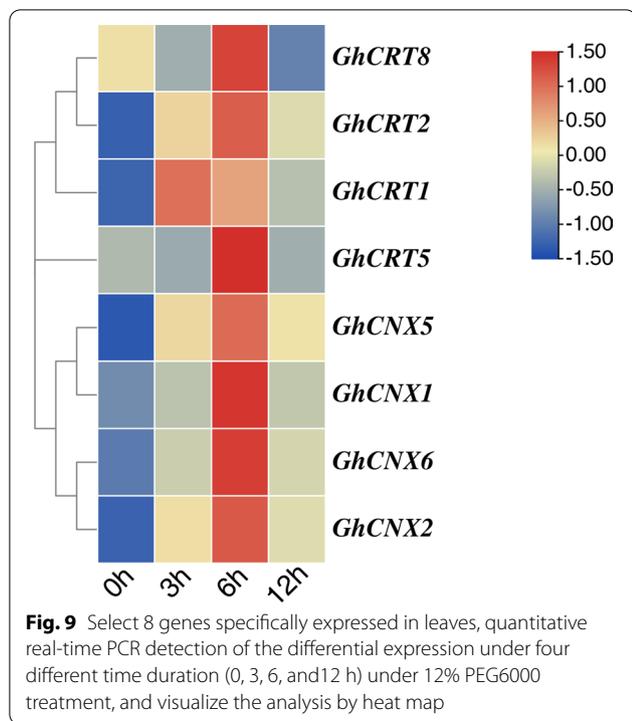
By constructing the co-expression network of *GhCNXs* family members, it was found that under the four kinds of abiotic stresses, family members played a common role in plant resistance to stress. According to the Pearson correlation coefficient (PCCs) co-expression network analysis of the relative expression of *GhCNXs* genes, multiple genes were correlated ( $P > 0.96$ ) under different abiotic stresses (cold, hot, salt, and PEG). Among them, 16 genes showed correlation under cold stress treatment (Fig. 10A), while 29 genes showed correlation under hot stress treatment (Fig. 10B). Also under salt stress, there were 31 pairs of genes that were negatively or positively



**Fig. 7** The specific expression of *GhCNX* genes in root, stem and in leaf. Significance level  $\alpha=0.05$ : The resulting values are expressed in relative units. The error bar in the figure is the standard deviation (SD) of the three biological replicates in each treatment group



**Fig. 8** The different expression patterns of *GhCNX* family members in *G. hirsutum* under stress. The leaf expression data of *GhCNX* family member genes under abiotic stress were obtained from *Gossypium* Resource And Network Database. The green bar graph shows the relative expression level of *CNX* family genes under 4°C. The orange bar graph shows the relative expression level of *CNX* family genes under 37°C. The blue bar graph represents the relative expression level of *CNX* family genes under PEG. The pink bar graph shows the relative expression level of *CNX* family genes under NaCl stress. CK was the control group; EG stands for experimental group. \*\**P* < 0.01; The resulting values are expressed in relative units. The error bar in the figure is the standard deviation (SD) of the three biological replicates in each treatment group



correlated (Fig. 10C). Sixty-four pairs of genes exhibited interactions under drought stress (Fig. 10D). Thus, co-expression network studies indicated that *GhCNXs* genes were highly affected by a variety of abiotic stresses. Among them, the correlation of *GhCNXs* gene is relatively close in response to PEG simulating natural drought stress (Fig. 10).

**Gossypium plants with GhCNX6 gene silenced by VIGS were sensitive to PEG**

By analyzing the differentially expressed genes of the *GhCNXs* family, a *GhCNX6* gene with high expression under different stresses was selected for further study. VIGS experiment was carried out on *GhCNX6* gene using cotton plant cv H177 as the material. After 2 weeks, PDS cotton plants appeared albino, indicating successful silencing of *GhCNX6*. The gene silencing effect was detected by qRT-PCR, and the results showed that the expression of pYL156:*GhCNX6* was significantly lower than that of the control pYL156. In summary, *GhCNX6* may play a role in the adaptation of cotton to PEG (Fig. 11).

**Discussion**

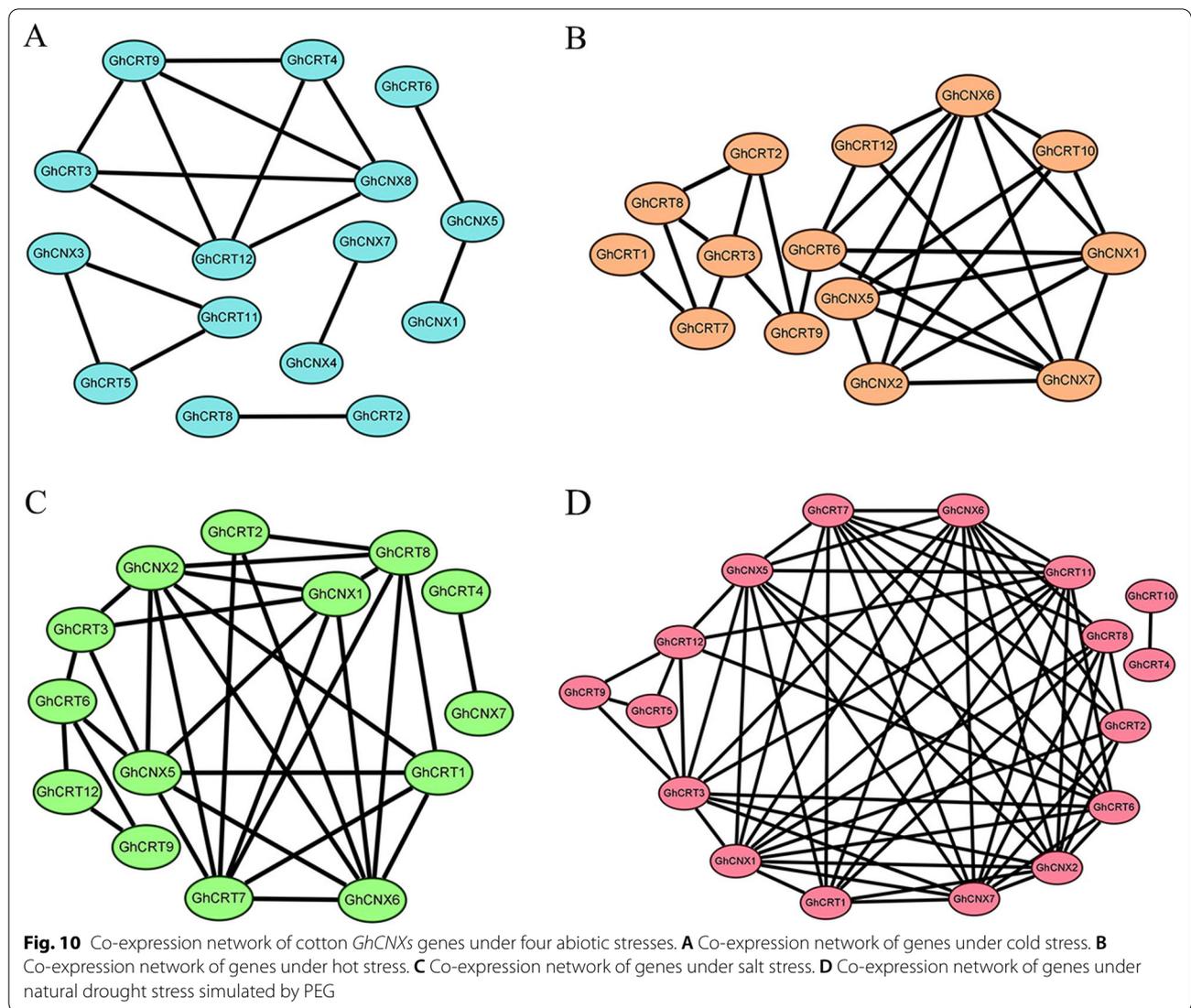
As an important economic crop widely grown, cotton faces severe biotic and abiotic stresses. Under stress, endoplasmic reticulum folding will be disturbed. *CNX* gene is an endoplasmic reticulum stress-related gene,

which plays a very important role in endoplasmic reticulum protein processing. Therefore, by analysing the phylogeny, gene structure and function of *CNX* family members, especially their expression patterns under abiotic stresses in *G. hirsutum*, these family members are closely related to cotton stress responses. As an evolutionarily conserved protein, *CNX* is mainly located in the ER. Although *CNX* was initially considered as a molecular chaperone to regulate the stability of  $Ca^{2+}$  internal environment, it also plays a role in many intracellular and extracellular growth and development (Wasag et al. 2019).

A total of 60 *CNX* family genes were identified in the four *Gossypium* species, it can be seen that they are highly conserved. By analyzing the physicochemical properties of the *CNX* family members of *G. hirsutum*, it is known that there are differences in the number of amino acid, molecular weight, isoelectric point and other aspects among the members, which may lead to differences in gene functions (Michu 2007). In the phylogenetic tree, all species were derived from the same node, suggesting that the *CNX/CRT* genes of these species all underwent gene duplication events that led to the expansion. However, gene duplication differs among species, and the *CNX* gene in cotton is relatively conserved but still shows evolutionary development. The numbers of genes in *G. hirsutum* and *G. barbadense* are 2~3 times compared with those in other species.

The distribution of genes on the chromosomes of *G. hirsutum* and *G. barbadense* is basically the same, and only a few genes are added or lost. The collinearity analysis of four *Gossypium* species showed that tandem duplication, segmental duplication, and whole genome duplication played a role in gene family expansion. The results showed that five tandem duplications with certain regularity occurred on chromosome 13. At the same time, 67 segmental duplications and 180 whole-genome replications were obtained. The evolution of *CNX* obviously involves gene replication events and new differentiation functions (An et al. 2011). After the selection pressure analysis, the gene family conservation was further verified, the results inferred that the *CNX* gene family has experienced strong purifying selection and limited functional divergence during the process of evolution (Hurst 2002).

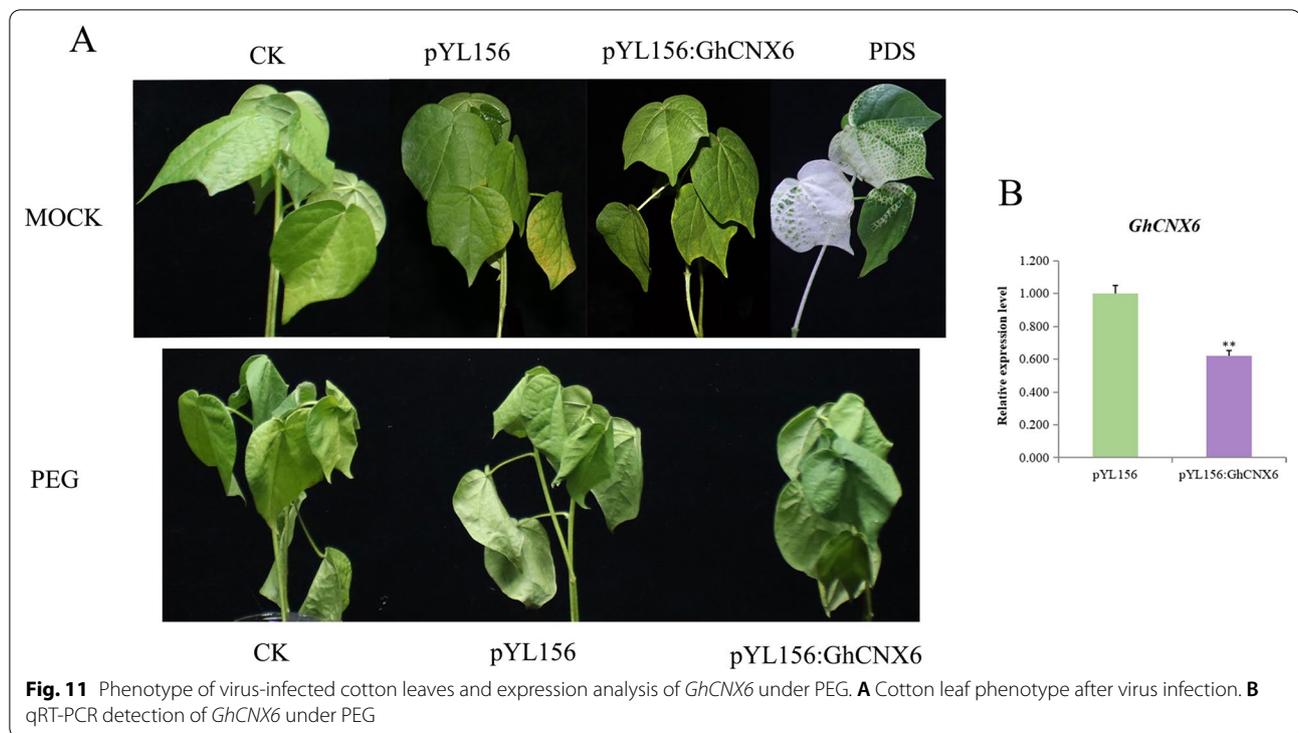
The analysis of *cis*-acting elements and differentially expressed genes shows that *CNX* family members is very important in the responding to various environmental stimuli. For example, some *CNX* or *CRT* genes from different duration time of PEG treatment shows obvious differences. By verifying the differentially expressed genes under different stresses, it was found that *GhCNX2*, *GhCNX5*, *GhCNX6*, *GhCRT1*, *GhCRT2*, and *GhCRT8*



were regulated by abiotic stress. The *cis*-acting elements of these gene promoter regions were predicted, including anaerobic, MeJA, abscisic acid, drought and involved in defence and stress responsiveness. It was speculated that some genes play an important role under biotic and abiotic stresses, which provides a reference for further functional verification. For example, previous studies have shown that plant *CRT* is involved in virus defence (Chen et al. 2005), plasmodesma, intra cellular transport (Laporte et al. 2003), ER calcium buffering, and stress response and tolerance (Jia et al. 2008).

In order to further verify the expression levels of *CNX* family genes under stress, quantitative real-time PCR of stem, root, and leaf tissues was performed with 20 genes. The expression levels of 20 genes were different among different tissues. It can be seen that

there are 8 genes with high expression levels in leaves. Then the eight genes were treated under abiotic stress and was found that the expression levels of different genes showed a certain expression trend after stress treatment. The expression of *GhCNX6* was down-regulated at 4°C and up-regulated at 37°C and PEG treatments. Then the expression of 8 genes in leaves treated with PEG was further analyzed. The expression levels were the highest after 6 h of PEG treatment, and then decreased after 12 h. The evolutionary relationship of the four genes *GhCNX1*, *GhCNX2*, *GhCNX5*, and *GhCNX6* is relatively close. At the same time, the expression levels of *GhCNX1* and *GhCNX6* tend to be the same after stress treatment. The expression levels of *GhCNX6* are relatively higher after 6 h under 12% PEG6000 treatment. It shows that the genes of this family have obvious differential expression patterns under



stress, which lays the foundation for the functional verification of some genes under stress conditions.

### Conclusions

In this study, the *CNX* gene family of four *Gossypium* species were comprehensively analyzed, and a total of 60 *CNX* family genes were identified. According to the analysis of phylogenetic tree, conserved domains and gene structure, the *CNX* genes of four *Gossypium* species were divided into three categories. Chromosomal positions and gene duplication relationship showed that the gene family expansion in cotton underwent through tandem, segmental and whole genome duplications. Expression analysis revealed the differential spatial expression patterns of *GhCNX* gene, indicating that *GhCNX* gene plays a vital role in regulating plant growth and development in the whole plant life cycle. By constructing co-expression networks under four abiotic stresses, it was found that the *CNX* family genes interacted together to cope with the stress environment. Especially under PEG treatment, the *CNX* family members act together through positive or negative regulation among genes. The results will provide an important basis for further in-depth study of the function of cotton *CNX* genes. However, the functional validation and biochemical characteristics of different members will provide clear clues for the specific roles of different types of *CNX* proteins.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42397-022-00133-8>.

- Additional file 1. Table S1:** *CNX* family genes corresponding to four *Gossypium* species genes renamed.
- Additional file 2. Table S2:** Analysis of the physical properties of *CNX* genes in *G. hirsutum*.
- Additional file 3. Table S3:** Primer pairs required for the qRT-PCR.
- Additional file 4. Table S4:** ID and protein sequences of 60 *CNX*s family members in cotton.
- Additional file 5. Table S5:** *CNX* family members *cis*-acting elements and their locations.
- Additional file 6. Table S6:** Relative expression levels of *CNX* family genes after four different abiotic stress treatments.
- Additional file 7. Table S7:** Relative expression levels of *CNX* family members in upland cotton tissues.
- Additional file 8. Table S8:** Relative expression levels of 8 selected genes under four abiotic stresses.
- Additional file 9. Table S9:** Relative expression levels of 8 selected genes under four different time treatments of 12% PEG6000.

### Acknowledgements

Not applicable.

### Author contributions

Conceived and designed the experiments: Ye WW, Xu N; methodology: Xu N, Zhang YX, Lu XK, Chen XG, Wang DL; experiment: Xu N, Zhang H, Fan YP, Rui C; analysis of data: Xu N, Zhang H, Wang J, Wang JJ, writing-original draft preparation: Xu N; writing-review and editing: Xu N, Malik W A, Han MG, Wang S, Chen C, Guo LX, Zhao LJ; supervision, Ye WW. All authors have read and approved the final manuscript.

### Funding

This work was supported by the Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences, and Supported by China Agriculture Research System of MOF and MARA.

### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no known competing interests or personal relationships.

Received: 30 June 2022 Accepted: 20 October 2022

Published online: 05 November 2022

### References

- An YQ, Lin RM, Wang FT, et al. Molecular cloning of a new wheat calreticulin gene *TaCRT1* and expression analysis in plant defense responses and abiotic stress resistance. *Genet Mol Res.* 2011;10:3576–85. <https://doi.org/10.4238/2011.November.10.1>.
- Chen CJ, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13(8):1194–202. <https://doi.org/10.1016/j.molp.2020.06.009>.
- Chen MH, Tian GW, Gafni Y, et al. Effects of calreticulin on viral cell-to-cell movement. *Plant Physiol.* 2005;138:1866–76. <https://doi.org/10.1104/pp.105.064386>.
- Del Bem LE. The evolutionary history of calreticulin and calnexin genes in green plants. *Genetica.* 2011;139:255–9. <https://doi.org/10.1007/s10709-010-9544-y>.
- Fan W, Zhang Z, Zhang Y. Cloning and molecular characterization of fructose-1,6-bisphosphate aldolase gene regulated by high-salinity and drought in *Sesuvium portulacastrum*. *Plant Cell Rep.* 2009;28:975–84. <https://doi.org/10.1007/s00299-009-0702-6>.
- Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. *The proteomics protocols handbook*. Springer protocols handbooks. Totowa, NJ: Humana Press. 2005; 571–607. <https://doi.org/10.1385/1-59259-890-0:571>.
- Guérin R, Arseneault G, Dumont S, et al. Calnexin is involved in apoptosis induced by endoplasmic reticulum stress in the fission yeast. *Mol Biol Cell.* 2008;19:4404–20.
- Hu Y, Chen J, Fang L, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet.* 2019;51:739–48. <https://doi.org/10.1038/s41588-019-0371-5>.
- Hurst LD. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. *Trends Genet.* 2002;18:486. [https://doi.org/10.1016/s0168-9525\(02\)02722-1](https://doi.org/10.1016/s0168-9525(02)02722-1).
- Hwang J, Qi L. Quality control in the endoplasmic reticulum: crosstalk between ERAD and UPR pathways. *Trends Biochem Sci.* 2018;43:593–605. <https://doi.org/10.1016/j.tibs.2018.06.005>.
- Jia XY, Xu CY, Jing RL, et al. Molecular cloning and characterization of wheat calreticulin (*CRT*) gene involved in drought-stressed responses. *J Exp Bot.* 2008;59:739–51. <https://doi.org/10.1093/jxb/erm369>.
- Krzywinski M, Schein J, Biro I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19:1639–45. <https://doi.org/10.1101/gr.092759.109>.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33:1870–4. <https://doi.org/10.1093/molbev/msw054>.
- Laporte C, Vetter G, Loudes AM, et al. Involvement of the secretory pathway and the cytoskeleton in intracellular targeting and tubule assembly of *Grapevine fanleaf virus* movement protein in tobacco BY-2 cells. *Plant Cell.* 2003;15:2058–75. <https://doi.org/10.1105/tpc.013896>.
- Li J, Zhang Z, Vang S, et al. Correlation between *Ka/Ks* and *Ks* is related to substitution model and evolutionary lineage. *J Mol Evol.* 2009;68:414–23. <https://doi.org/10.1007/s00239-009-9222-9>.
- Liu DY, Smith PM, Barton DA, et al. Characterisation of *Arabidopsis* calnexin 1 and calnexin 2 in the endoplasmic reticulum and at plasmodesmata. *Protoplasma.* 2017;254:125–36. <https://doi.org/10.1007/s00709-015-0921-3>.
- Malik WA, Wang X, Wang X, et al. Genome-wide expression analysis suggests glutaredoxin genes response to various stresses in cotton. *Int J Biol Macromol.* 2020;153:470–91. <https://doi.org/10.1016/j.ijbiomac.2020.03.021>.
- Michu E. A short guide to phylogeny reconstruction. *Plant Soil Environ.* 2007;53:442–6. <https://doi.org/10.17221/2194-Pse>.
- Nakao H, Seko A, Ito Y, et al. PDI family protein ERp29 recognizes P-domain of molecular chaperone calnexin. *Biochem Biophys Res Commun.* 2017;487:763–7. <https://doi.org/10.1016/j.bbrc.2017.04.139>.
- Nguyen VC, Nakamura Y, Kanehara K. Membrane lipid polyunsaturation mediated by FATTY ACID DESATURASE 2 (FAD2) is involved in endoplasmic reticulum stress tolerance in *Arabidopsis thaliana*. *Plant J.* 2019;99:478–93. <https://doi.org/10.1111/tpj.14338>.
- Nouri MZ, Hiraga S, Yanagawa Y, et al. Characterization of calnexin in soybean roots and hypocotyls under osmotic stress. *Phytochemistry.* 2012;74:20–9. <https://doi.org/10.1016/j.phytochem.2011.11.005>.
- Oelze ML, Muthuramalingam M, Vogel MO, et al. The link between transcript regulation and de novo protein synthesis in the retrograde high light acclimation response of *Arabidopsis thaliana*. *BMC Genomics.* 2014;15:320. <https://doi.org/10.1186/1471-2164-15-320>.
- Parlati F, Dominguez M, Bergeron JJ, et al. *Saccharomyces cerevisiae* CNE1 encodes an endoplasmic reticulum (ER) membrane protein with sequence similarity to calnexin and calreticulin and functions as a constituent of the ER quality control apparatus. *J Biol Chem.* 1995;270:244–53. <https://doi.org/10.1074/jbc.270.1.244>.
- Prince VE, Pickett FB. Splitting pairs: the diverging fates of duplicated genes. *Nat Rev Genet.* 2002;3:827–37. <https://doi.org/10.1038/nrg928>.
- Rao RV, Bredesen DE. Misfolded proteins, endoplasmic reticulum stress and neurodegeneration. *Curr Opin Cell Biol.* 2004;16:653–62. <https://doi.org/10.1016/j.ceb.2004.09.012>.
- Ruggiano A, Foresti O, Carvalho P. Quality control: ER-associated degradation: protein quality control and beyond. *J Cell Biol.* 2014;204:869–79. <https://doi.org/10.1083/jcb.201312042>.
- Sakono M, Seko A, Takeda Y, et al. Glycan specificity of a testis-specific lectin chaperone calmeglin and effects of hydrophobic interactions. *Biochim Biophys Acta.* 2014;1840:2904–13. <https://doi.org/10.1016/j.bbagen.2014.04.012>.
- Schrag JD, Procopio DO, Cygler M, et al. Lectin control of protein folding and sorting in the secretory pathway. *Trends Biochem Sci.* 2003;28:49–57. [https://doi.org/10.1016/s0968-0004\(02\)00004-x](https://doi.org/10.1016/s0968-0004(02)00004-x).
- Sitia R, Braakman I. Quality control in the endoplasmic reticulum protein factory. *Nature.* 2003;426:891–4. <https://doi.org/10.1038/nature02262>.
- Valente MA, Faria JA, Soares-Ramos JR, et al. The ER luminal binding protein (BiP) mediates an increase in drought tolerance in soybean and delays drought-induced leaf senescence in soybean and tobacco. *J Exp Bot.* 2009;60:533–46. <https://doi.org/10.1093/jxb/ern296>.
- Vu KV, Nguyen NT, Jeong CY, et al. Systematic deletion of the ER lectin chaperone genes reveals their roles in vegetative growth and male gametophyte development in *Arabidopsis*. *Plant J.* 2017;89:972–83. <https://doi.org/10.1111/tpj.13435>.
- Wang D, Zhang S, He F, et al. How do variable substitution rates influence *Ka* and *Ks* calculations? *Genomics Proteomics Bioinform.* 2009;7:116–27.
- Wang XG, Lu XK, Malik WA, et al. Differentially expressed bZIP transcription factors confer multi-tolerances in *Gossypium hirsutum* L. *Int J Biol Macromol.* 2020;146:569–78. <https://doi.org/10.1016/j.ijbiomac.2020.01.013>.
- Wang Y, Tang H, Debarry JD, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40:e49. <https://doi.org/10.1093/nar/gkr1293>.
- Wasag P, Grajkowski T, Suwinska A, et al. Phylogenetic analysis of plant calreticulin homologs. *Mol Phylogenet Evol.* 2019;134:99–110. <https://doi.org/10.1016/j.jmpev.2019.01.014>.

- Xu GX, Guo CC, Shan HY, et al. Divergence of duplicate genes in exon-intron structure. *Proc Natl Acad Sci U S A*. 2012;109:1187–92. <https://doi.org/10.1073/pnas.1109047109>.
- Zhang Z, Chai M, Yang Z, et al. GRAND: an integrated genome, transcriptome resources, and gene network database for *Gossypium*. *Front Plant Sci*. 2022;13:773107. <https://doi.org/10.3389/fpls.2022.773107>.
- Zhu T, Liang CZ, Meng ZG, et al. CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol*. 2017;17:101. <https://doi.org/10.1186/s12870-017-1039-x>.
- Zhu W, Tan W, Li Q, et al. Genome-wide characterization and expression profiling of the *MAPKKK* genes in *Gossypium arboreum* L. *Genome*. 2019;62:609–22. <https://doi.org/10.1139/gen-2018-0176>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

