Genome-wide identification and characterization of the CLASP_N gene family in upland cotton (*Gossypium hirsutum* L.)

Meijun Ji[1,*], Kangtai Sun[1,*], Hui Fang[1], Zhimin Zhuang[1], Haodong Chen[2], Qi Chen[1], Ziyi Cao[1], Yiting Wang[1], Allah Ditta[3], Muhammad Kashif Riaz Khan[3], Kai Wang[1] and Baohua Wang[1]

[1] School of Life Sciences, Nantong University, Nantong, Jiangsu, China
[2] Cotton Sciences Research Institute of Hunan/ National Hybrid Cotton Research Promotion Center, Changde, Hunan, China
[3] Plant Breeding and Genetics Division, Nuclear Institute for Agriculture and Biology, Faisalabad, Pakistan
[*] These authors contributed equally to this work.

ABSTRACT

Background. Cytoplasmic linker–associated proteins (CLASPs) are tubule proteins that can bind to microtubules and participate in regulating the structure and function of microtubules, which significantly affects the development and growth of plants. These proteins have been identified in *Arabidopsis*; however, little research has been performed in upland cotton.

Methods. In this study, the whole genome of the CLASP_N family was analyzed to provide theoretical support for the function of this gene family in the development of upland cotton fiber. Bioinformatics was used to analyze the family characteristics of CLASP_N in upland cotton, such as member identification, sequence characteristics, conserved domain structure and coevolutionary relationships. Real-time fluorescent quantitative PCR (qRT-PCR) was used to clarify the expression pattern of the upland cotton CLASP_N gene family in cotton fiber.

Results. At the genome-wide level, we identified 16 upland cotton CLASP_N genes. A chromosomal localization analysis revealed that these 16 genes were located on 13 chromosomes. The motif results showed that all CLASP_N proteins have the CLASP_N domain. Gene structure analysis showed that the structure and length of exons and introns were consistent in the subgroups. In the evolutionary analysis with other species, the gene family clearly diverged from the other species in the evolutionary process. A promoter sequence analysis showed that this gene family contains a large number of cis-acting elements related to a variety of plant hormones. qRT-PCR was used to clarify the expression pattern of the upland cotton CLASP_N gene family in cotton fiber and leaves, and *Gh210800* was found to be highly expressed in the later stages of fiber development. The results of this study provide a foundation for further research on the molecular role of the CLASP_N genes in cotton fiber development.

Subjects Agricultural Science, Bioinformatics, Genomics, Molecular Biology, Plant Science
Keywords *Gossypium hirsutum*, CLASP_N, Gene family, Fiber development

INTRODUCTION

Upland cotton is the most important natural fiber crop in the world (*Chen et al., 2007*). There are four cultivated species, including the two tetraploid species *Gossypium hirsutum* (upland cotton) and *G. barbadense* (sea island cotton) and the two diploid cotton species *G. herbaceum* and *G. arboreum* (*Alkuddsi et al., 2013*). Upland cotton accounts for more than 95% of the world's cotton fiber production (*Song et al., 2019*). In contrast, sea island cotton (*G. barbadense*) produces higher-quality fiber but lower yields (*Shi et al., 2015*). Although fiber quality has been greatly improved in recent decades by conventional cultivar breeding programs, the demand for higher fiber quality in upland cotton has increased with the development of the modern textile industry and the need for diverse upland cotton products. The fiber length (FL), fiber strength (FS), fiber length uniformity (FU), fiber elongation (FE), and micronaire value (MIC) are important indicators for evaluating fiber quality (*Wang et al., 2016a*; *Wang et al., 2016b*), while maturity and fiber length are important factors that affect spinning (*Kelly et al., 2019*). As single-celled trichomes formed by ovule epidermal cells, cotton fibers are outstanding model systems to study plant cell elongation and cell wall and cellulose biosynthesis (*Kim & Triplett, 2001*). Approximately one-third of epidermal cells differentiate into spinnable fibers. The development process of cotton fiber includes the primary stage, elongation stage, secondary wall biosynthesis stage and mature stage. These four stages are distinct but overlapping (*Kim & Triplett, 2001*). In the elongation phase, fiber cells rapidly elongate and gradually reach their maximum length (*Fang et al., 2014*). In cultivated varieties of cotton, seed trichomes are composed of long lint fibers that are easily detached from the seed and fuzz fibers that are tightly attached to the seed (*Applequist, Cronn & Wendel, 2001*). In addition, fuzz fibers cannot be twisted many times during spinning. Therefore, long fibers should be used because they are easier to twist. Cotton fiber is a highly elongated and thickened single cell. The growth of cotton fiber cells undergoes stage differentiation, including the synthesis of primary wall in the elongation stage and the synthesis of cellulose in the thickening stage of the secondary wall. It is worth noting that the deposition of high ratio of cellulose in mature cotton fibers makes the fibers have good hardness and strength. This process is of great significance in nature and industry (*Tang & Liu, 2017*). Therefore, the elongation period and the thickening of the secondary cell wall are critical periods for the development of cotton fibers, which are worthy of further study.

In eukaryotes, microtubules and microtubule-associated proteins control cell growth and development by organizing microtubule arrays and spatial aspects (*Hamada, 2014*). Studies have shown that CLASPs mediate the function of *Arabidopsis* microtubules in response to mechanical stress on different spatial scales. This protein responds to different degrees of tissue-scale mechanical stress in the cotyledon epidermis by regulating the organization of microtubules (*Eng et al., 2021*). At present, many types of microtubule binding protein families have been discovered. The microtubule-binding proteins of these families affect microtubule organization through various functions, including regulating the microtubule dynamic properties, microtubule interactions, and microtubule cleavage

(*Kong et al., 2015*). Cytoplasmic linker–associated proteins (CLASPs) are microtubule-binding proteins that can regulate the structure and function of microtubules (*Pietra et al., 2013*) and bind to microtubules and cytoplasmic connexins. Because it binds to the positive end of microtubules, CLASP is also called a positive tracer protein (*Cleavage Akhmanova & Hoogenraad, 2005*). This protein contains a typical conserved domain of the CLASP_N terminus, and some of the proteins also contain HEAT repeat domains (*Sedbrook, 2004*). The CLASP_N terminus usually contains two TOG or TOG-like (TOGL) domains and a basic serine-rich domain (*Caroline et al., 2014*). In eukaryotes, this protein regulates microtubule dynamic activity through these functional domains. There are few reports on the function of CLASP in plants, and most are based on the model plant *Arabidopsis thaliana*. There is only one CLASP gene in the *Arabidopsis* genome (*Ambrose et al., 2007*). Microtubules are the material transport tracks in cells, and damage to microtubules will hinder the material transport in cells (*Gardiner, 2013*). The arrangement of periplasmic microtubules is related to fiber cell elongation and secondary wall biosynthesis (*Sumino et al., 2012*). Studies have shown that the CLASP genes are related to the growth of root meristems, the elongation of hypocotyl cells, and leaf epidermal hair branching (*Laryssa, Katherine & Geoffrey, 2020*). Cotton fibers are single cells derived from elongated cells of the ovule epidermis. It is speculated that cotton fiber may have a similar regulatory mechanism to the development of the *Arabidopsis* epidermal hair with similar structural (*Le & Ambrose, 2018*). Since related genes have been identified in *Arabidopsis*, it is speculated that the CLASP_N gene family may play a role in the development of cotton fiber, *e.g.*, *Zhu et al. (2018)* found that *GhCLASP* is mainly expressed in stem fibers and developing fibers. Based on previous studies, we used the latest upland cotton reference genome to complete the genomic identification of the CLASP_N gene family. At the same time, we carried out chromosome mapping, multisequence alignment, and evolutionary tree construction for the genes of the family and motif prediction and gene structure identification of the CLASP_N gene family genes. This research will help reveal the molecular role of the *CLASP_N* genes in cotton fiber development.

MATERIALS & METHODS

Cotton material

The public upland cotton germplasm PD94042, which is characterized by improved fiber maturity and high yield potential (*May, 1999*), was used in this research. First, seeds with a full grain and similar sizes were selected. Seeds of PD94042 were planted in a field at Nantong University, and flowers were tagged before anthesis. The leaves of the one-leaf stage and the fibers at 17 and 21days post anthesis (DPA) were collected, and RNA was extracted for RT-qPCR validation.

Identification and sequence analysis of CLASP_N family members in upland cotton

Whole genome data of upland cotton were downloaded from the JGI plant genome database (https://phytozome-next.jgi.doe.gov/info/Ghirsutum_v2_1) (*Chen et al., 2020*). The Hidden Markov Model (HMM) database (Protein families database of alignments)

and Pfam database (http://pfam.xfam.org/) (*Finn et al., 2016*) were used to download the seed file PF12348 of the *CLASP_N* genes. After the screening, the protein sequences were uploaded to the online software Pfam, the Simple Modular Architecture Research Tool database (SMART, http://smart.embl-heidelberg.de) (*Letunic, 2017*) and the Conserved Domain database (CDD, https://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), and sequence alignment and analysis were carried out to remove unannotated genes and redundant sequences (*Marchler-Bauer et al., 2015*). The gene IDs containing the CLASP_N special domain and its corresponding sequence were obtained. The functional annotation and GO annotation information of each gene was searched through the websites of https://jgi.doe.gov/ and http://geneontology.org/ respectively.

Chromosome location and collinearity analysis

The gff3 annotation file of the genome of upland cotton was used to extract biological information, such as the position and structure of CLASP_N family members. Mapchart software (*Voorrips, 2002*) was used to analyze and demonstrate the chromosomal positions of the family genes. AI (Adobe Illustrator CS6) software was used to beautify the chromosome position map. To reveal the collinearity of the CLASP_N family among upland cotton species, all protein sequences in upland cotton were BLAST searched and compared using MCSCANX software (*Wang et al., 2012*), and collinear genes were found in the whole genome. The collinearity of genes in the gene family was disclosed, and the gene family circle was drawn using MCSCANX software with special annotation.

Gene structure and motif analysis

The location information of all *CLASP_N* genes, including the exons, introns, and UTR location information on the chromosome, was available in the reference genome of the upland cotton database. A gene structure diagram (*Hu et al., 2015*) was drawn using the online website GSDS (http://gsds.cbi.pku.edu.cn/). MEME software (http://meme-suite.org/) (*Bailey et al., 2006*) was applied to carry out a motif analysis based on the protein sequences of *CLASP_N* genes. The exon-intron structure of the gene family was analyzed with the online software GSDS (http://gsds.cbi.pku.edu.cn/). TBTools was used to combine and visualize the analyzed sequence alignment results files, exon-intron structure files, and conservative domain files (*Zhang, Raikhel & Hicks, 2013*).

Construction of phylogenetic tree of CLASP_N family members

For the identified *CLASP_N* gene members, the protein sequences of upland cotton, *Arabidopsis thaliana*, rice (*Oryza sativa*), tomato (*Solanum lycopersicum*), and corn (*Zea mays*) were extracted using HMMER 3.0 and a BLASTP search. The extracted protein sequences were then compared with MEGA 7.0 software (*Sudhir, Glen & Koichiro, 2016*). A phylogenetic tree was then built based on the comparison results. The evolutionary tree was constructed with neighbor-joining method (*Li et al., 2019*). The Bootstrap repeat test (Replications) was set to 1,000. The online software Evolview (http://www.omicsclass.com/article/671) was used to beautify the evolutionary tree.

Cis-elements analysis of *CLASP_N* genes

To conduct the functional correlation analysis, 1.5 kb sequences located upstream of ATG of all *CLASP_N* genes were extracted to analyze the cis-elements of genes. The PlantCare (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) database (*Lescot et al., 2002*) was used to recognize and analyze the cis-acting elements of genes, and then they were visualized using TBTools.

RNA extraction and quantitative real-time PCR (qRT-PCR) analysis

Total RNA was extracted from cotton fiber samples using a polysaccharide polyphenol complex plant RNA extraction kit (AIDLAB, Beijing, China) following the manufacturer's procedure. A maximum of 1 µg total RNA was used to synthesize cDNA by HiScript QRT SuperMix (TAKARA, Beijing, China), and then qRT-PCR was performed with an Opticon thermocycler (CFXConnect Real-Time System; Bio-Rad, Hercules, CA) using SYBR Green PCR master mix (TAKARA, Beijing, China) according to the manufacturer's instructions. Primer 5 was used to design the relevant primers needed for the PCR assay (Table S1). The PCR conditions were set as follows: 95 °C for 10 min, 40 cycles of 95 °C for 10 s and 60 °C for 30 s; and a 65–95 °C melt curve, which was analyzed to detect possible nonspecific amplification or primer dimers. The upland cotton histidine gene was used as a control for normalization between samples. The comparative threshold cycle method was used to calculate the relative transcript levels (*Ma & Zhao, 2010*).

RESULTS

Genome-wide identification of CLASP_N family members

Using domain information (PF12348) obtained from domain prediction to identify members of the whole upland cotton genome, 35 genes were obtained. After comparing the results from the CDD, SMART, and other databases, a total of 16 genes were finally identified as belonging to the CLASP_N gene family. The gene information of upland cotton was searched by HMM and shown in Table S2. To maintain consistency, all upland cotton genes were temporarily named *Gh147000-Gh053074* according to the gene IDs. Functional annotation showed that most genes had similar functions, which were mainly related to spiral and folding. GO annotation indicated that the biological process involved in genes *Gh47000, Gh142700, Gh308900, Gh152100, Gh140100* and *Gh309600* are the regulation of lipid kinase activity, and the cellular components involved are spindle microtubules.

Chromosomal localization and collinearity analysis in CLASP_N

Chromosome positioning was carried out with the *CLASP_N* genes (Fig. 1). Our results indicated that the distribution of genes was uneven, with some chromosomes showing an overdistributed or tight arrangement in certain areas. Among them, 16 *CLASP_N* genes of upland cotton were located on 13 chromosomes and distributed on A01, A04, A05, A07, A08, A09, A12, D01, D04, D05, D07, D08 and D09. Most of them were located at the distal or proximal ends of the chromosomes.

To understand the evolutionary relationship of *CLASP_N* genes, collinearity analysis was conducted on *CLASP_N* genes of tetraploid upland cotton (Fig. 2), and there were 12
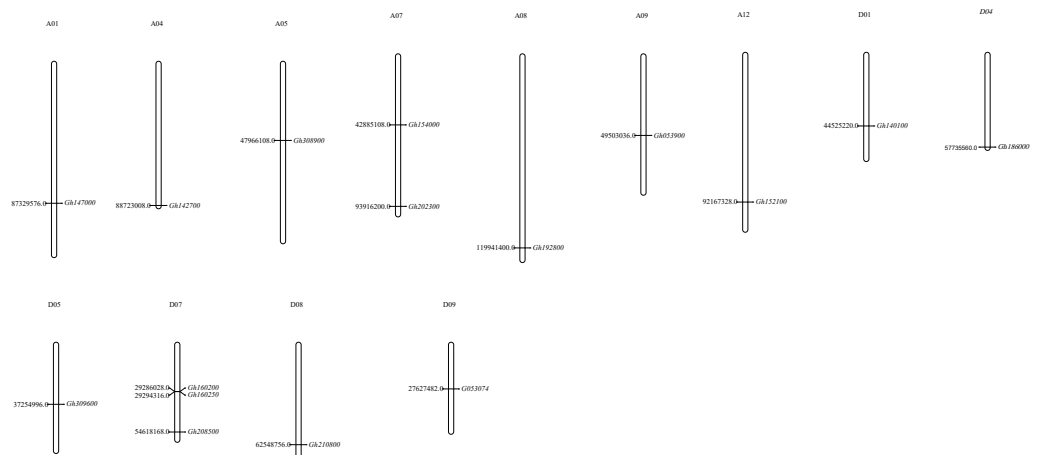
Figure 1 Chromosomal distribution of *CLASP_N* genes in upland cotton.

homologous gene pairs (Table S3). According to our study, there were many homologous gene pairs between D07 and D08, which indicated that different subgroups of upland cotton share a close evolutionary relationship. These results indicate that the gene family has undergone genomic rearrangement during the process of polyploidization.

Multiple sequence alignment and gene structure analysis

The structures of 16 *CLASP_N* genes in upland cotton were detected by the GSDS website to further understand the evolution of the gene structure of the CLASP_N family in upland cotton. First, the exon-intron structure of the 16 *CLASP_N* genes in upland cotton was studied, and we found that the number of exons of *CLASP_N* genes in upland cotton was larger than that of introns. Among them, eight genes had 21 exons, seven genes had 22 exons, and only the *Gh160250* gene had nine exons. Most of the exons of *CLASP_N* genes were densely distributed.

The analysis of the conserved motif of CLASP_N in upland cotton by MEME software (Fig. 3) showed that motif 2, motif 7, and motif 8 were all contained in the 6 subgroups of upland cotton, indicating that these 3 motifs were the most conserved motifs in the CLASP_N family. Motif 9 is unique to one group, while motif 4 is unique to two or three groups. However, the function of these conserved motifs remains to be studied. Generally, the gene structure analysis indicated that clusters in the same class presented relatively strong gene structure consistency. Moreover, the CLASPs of upland cotton in the same group usually contained similar motifs, indicating that they might play similar roles in the growth and development of upland cotton.

Phylogenetic analysis of upland cotton CLASP_N

To understand the evolutionary relationship of the CLASP_N family, MEGA 7.0 software was used to compare and analyze the amino acid sequences of upland cotton (*Chen et al., 2020*), *Arabidopsis thaliana* (*Cheng et al., 2017*), tomato (*Solanum lycopersicum*) (*Aflitos et al., 2015*), rice (*Oryza sativa*) (*Kawahara et al., 2013*), and corn (*Zea mays*)
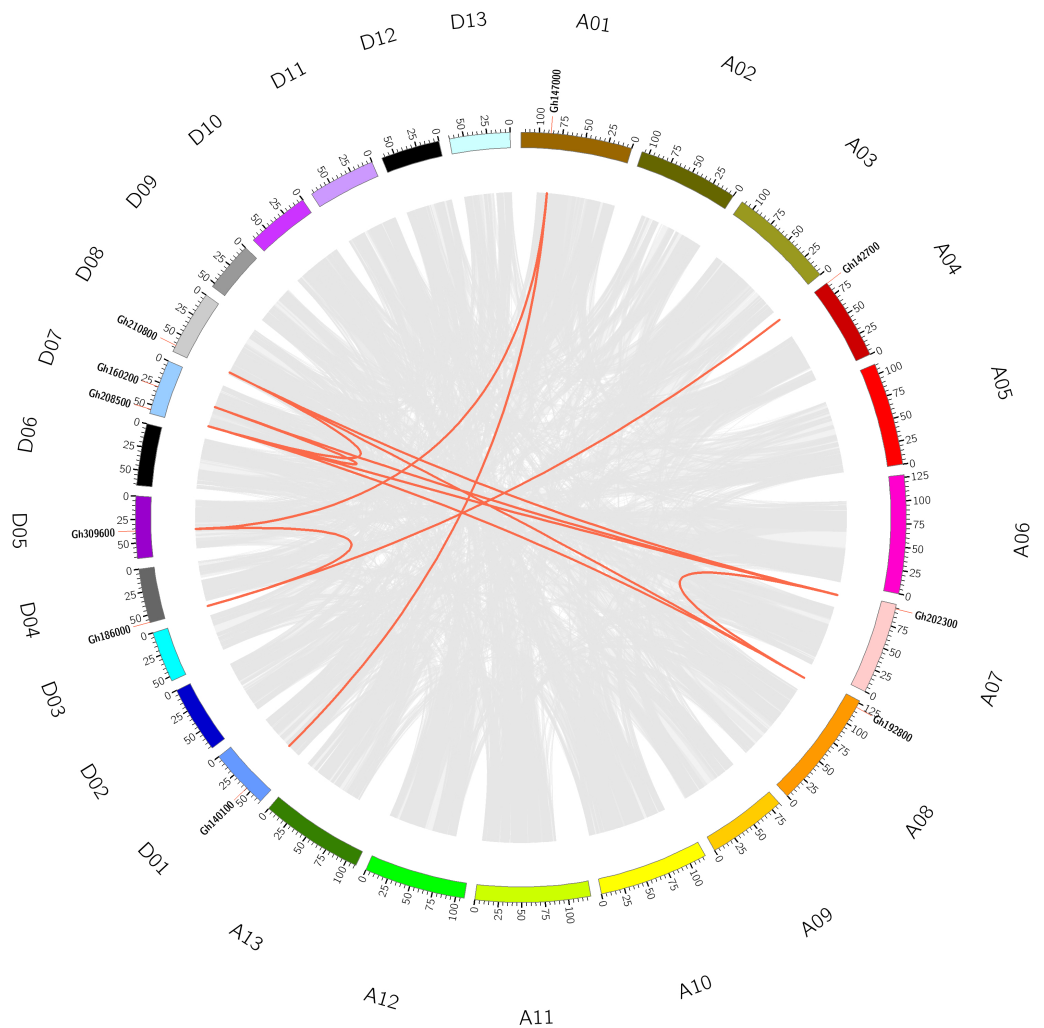
Figure 2 **The collinearity of *CLASP_N* genes in the upland cotton.** The gray lines represent collinear relationships within different genomes, and the red lines represent collinear gene pairs in the CLASP_N family.

(*Wang et al., 2016a*; *Wang et al., 2016b*) using the neighbor-joining method, and a phylogenetic tree was constructed (Fig. 4). According to our analysis and comparison results, the gene family is divided into three major groups, Group 1, Group 2 and Group 3, and each group is divided into several subfamilies. The results showed that the CLASP_N family members of upland cotton only existed in Group 1 and Group 2. Group 1 contained fewer family members and only contained family members of upland cotton. Genes from the same subgroup can be considered to have the same function. Group 2 and Group 3 contained almost the same number of family members. These results indicate genetic differentiation between *CLASP_N* genes in other species and upland cotton.

In addition, the upland cotton genes in Group 1 contained fewer motif types but were similar in type. The members of the upland cotton family in Group 2 contained a
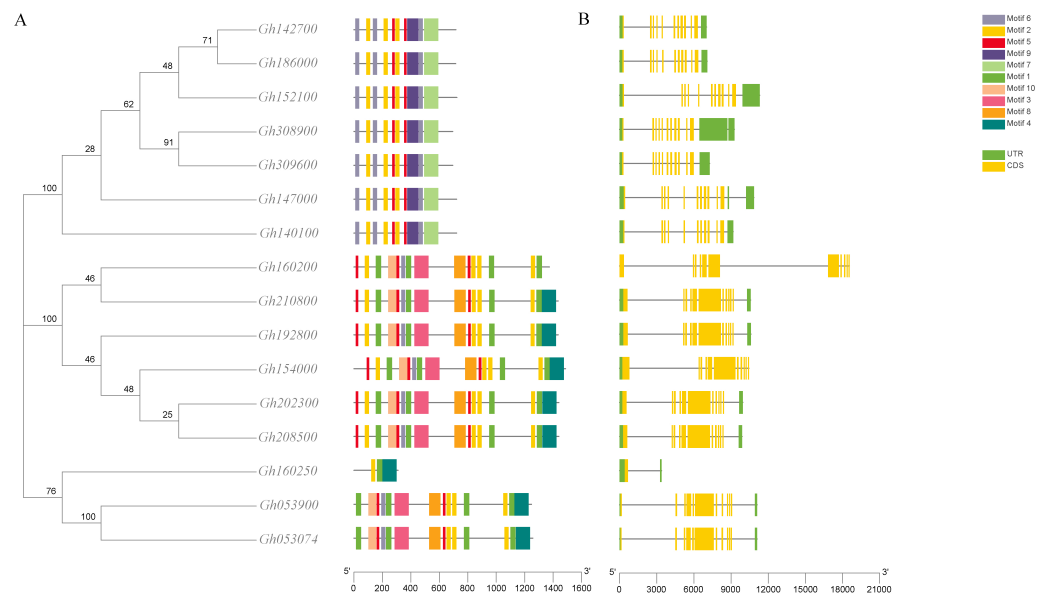
Figure 3 **Phylogenetic tree, conserved motif and gene structure of CLASP_N protein in upland cotton.**
(A) Motif distribution of CLASP_N proteins. The conserved motifs in the CLASP_N proteins were identified with MEME software. Grey lines denote the non-conserved sequences, and each motif is indicated by a colored box numbered at the bottom. The length of motifs in each protein was presented proportionally. (B) The exon-intron structure of CLASP_N genes based on the evolutionary relationship. The yellow rectangle represents exons, the grey line represents introns, and the green rectangle represents UTR.
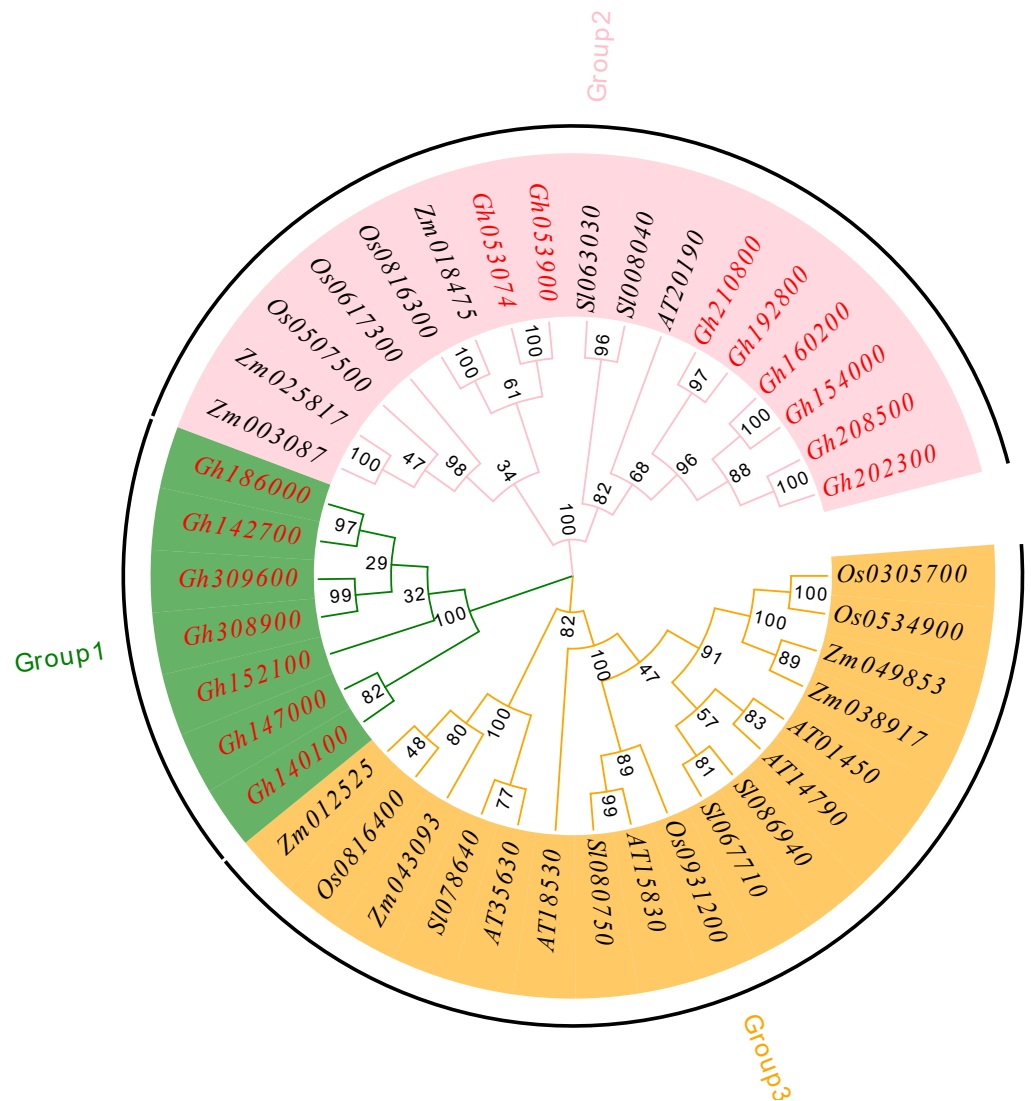
Full-size 🖼 DOI: 10.7717/peerj.12733/fig-3

larger number of motifs, and the types of motifs contained in each member were generally similar. The results showed that all *CLASP_N* genes in the same subfamily had conserved motifs and similar gene structures, which could support the reliability of the phylogenetic classification.

Cis-element analysis of promoter sequence

To further study the regulatory mechanism of the upland cotton CLASP_N gene family in the abiotic stress response, 16 gene family sequences were extracted from the upland cotton genome information for a cis-acting element analysis (Table S4), and the results showed that 16 abiotic stress response elements were involved (Fig. 5). In addition to a large number of basic elements, CAAT boxes and TATA boxes, in the CLASP_N gene family of upland cotton, there are also cis-acting regulatory elements involved in the light response, such as G-box and light response (*Zhang et al., 2020*). The MYB binding site of the module part is a WRE3 element that participates in the photoreaction. MYB participates in the regulation of plant phenylpropane secondary metabolic pathways (*Arce-Rodríguez, Martínez & Ochoa-Alejo, 2021*). The WUN motif is a stress response element in plants (*He et al., 2021*), and ARE is a cis-acting regulatory element necessary for anaerobic induction (*Mineo et al., 2021*). The results of the study indicate that members of the CLASP_N gene family of upland cotton are involved in the process of responding to a variety of biotic and abiotic stresses.

Figure 4 Phylogenetic tree of *CLASP_N* gene family members.

Full-size ⬚ DOI: 10.7717/peerj.12733/fig-4

Expression pattern of *CLASP_N* genes at different times

To further explore the expression changes in the *CLASP_N* genes in different tissues at different time, qRT-PCR was used to investigate the transcript levels of each *CLASP_N* gene at leaves, fibers of 17 DPA and 21 DPA, respectively, and we performed 3 biological repetitions and 2 technical repetitions (Table S5). Generally, most of the expression levels at fiber-21 DPA were lower than those at fiber-17 DPA (Fig. 6). The relative expression at fiber-21 DPA was significantly higher than at fiber-17 DPA only for *Gh210800*. The expression levels of fiber-21 DPA in *Gh152100*, *Gh192800*, *Gh147000*, *Gh309600*, *Gh160250*, *Gh140100*, and *Gh186000* were lower than that at one-leaf stage and fiber-17DPA. For *Gh154000*, *Gh202300*, *Gh208500*, and *Gh210800*, the expression level at both fiber stages was significantly higher than that at the one-leaf stage. The expression of *Gh210800* is
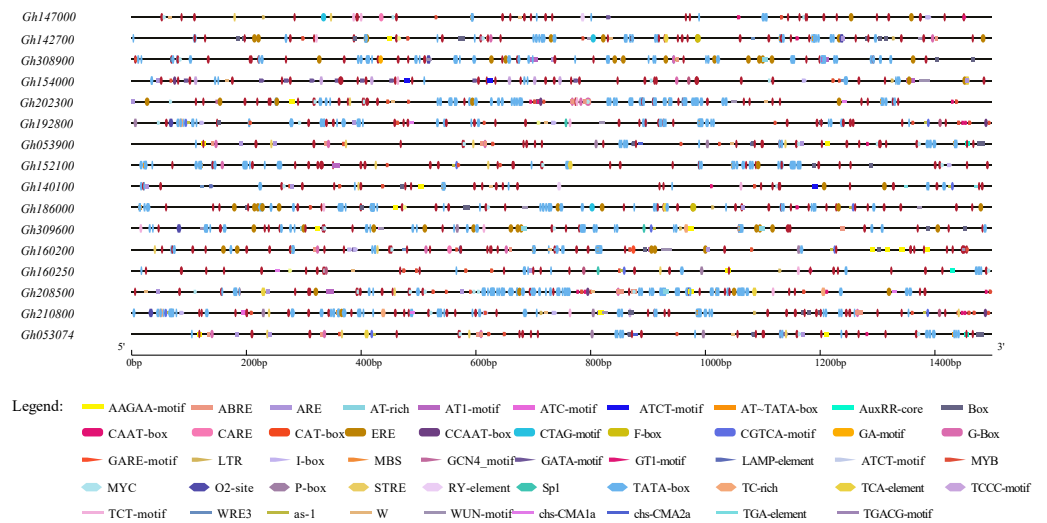
Figure 5 Homeopathic element analysis of upland cotton *CLASP_N* gene family.

Full-size DOI: 10.7717/peerj.12733/fig-5

significantly higher than that of the other genes, and its expression level increased with time past.

DISCUSSION

CLASPs are proteins that bind to the positive end of microtubules and have the ability to bind to both microtubule and cytoplasmic junction proteins. CLASPs generally contain two conserved domains: the HEAT repeat domain and the CLASP_N terminal domain. Some CLASPs contain only the CLASP_N structure. The N-terminus of most CLASPs contains two TOG or TOG-like (TOGL) domains, followed by a basic serine-rich domain. In vitro, this domain is sufficient for fission yeast Peg1/Cls1 to bind to the spatial regulation of microtubule (MT) lattice with high affinity (*Al-Bassam et al., 2010*), whereas Stu1, a CLASP essential for spindle integrity, additionally requires part of the TOGL2 domain (*Yin et al., 2002*). Stu1 in *Saccharomyces cerevisiae* is needed for the MTs in vivo. Depending on the stage of mitosis, Stu1 is also recruited to the spindle MTs through different mechanisms: in the middle stage, Stu1 directly binds to the MT lattice, while in the later stage, it is located indirectly in the middle of the spindle (*Caroline et al., 2014*). Similarly, in *S. pombe*, Cls1 localizes to the spindle midzone in anaphase (*Caroline et al., 2014*). In vivo and *in vitro* experiments have demonstrated that the HEAT domain can form helix structures, regulate the interactions between proteins and participate in the transport of intracellular substances (*Groves et al., 1999*). The CLASP_N terminal domain is a very conserved domain, which is the binding site of microtubules and has the stabilizing effect of regulating dynamic microtubules. In this study, a total of 16 members of the CLASP genes in upland cotton were identified. The proteins encoded by the genes varied in length, but all of them had the typical CLASP_N terminal domain. In vivo, CLASPs regulate only a subset of cellular MTs, which may be caused by specific CLASP localization (*Al-Bassam et al., 2010*). During
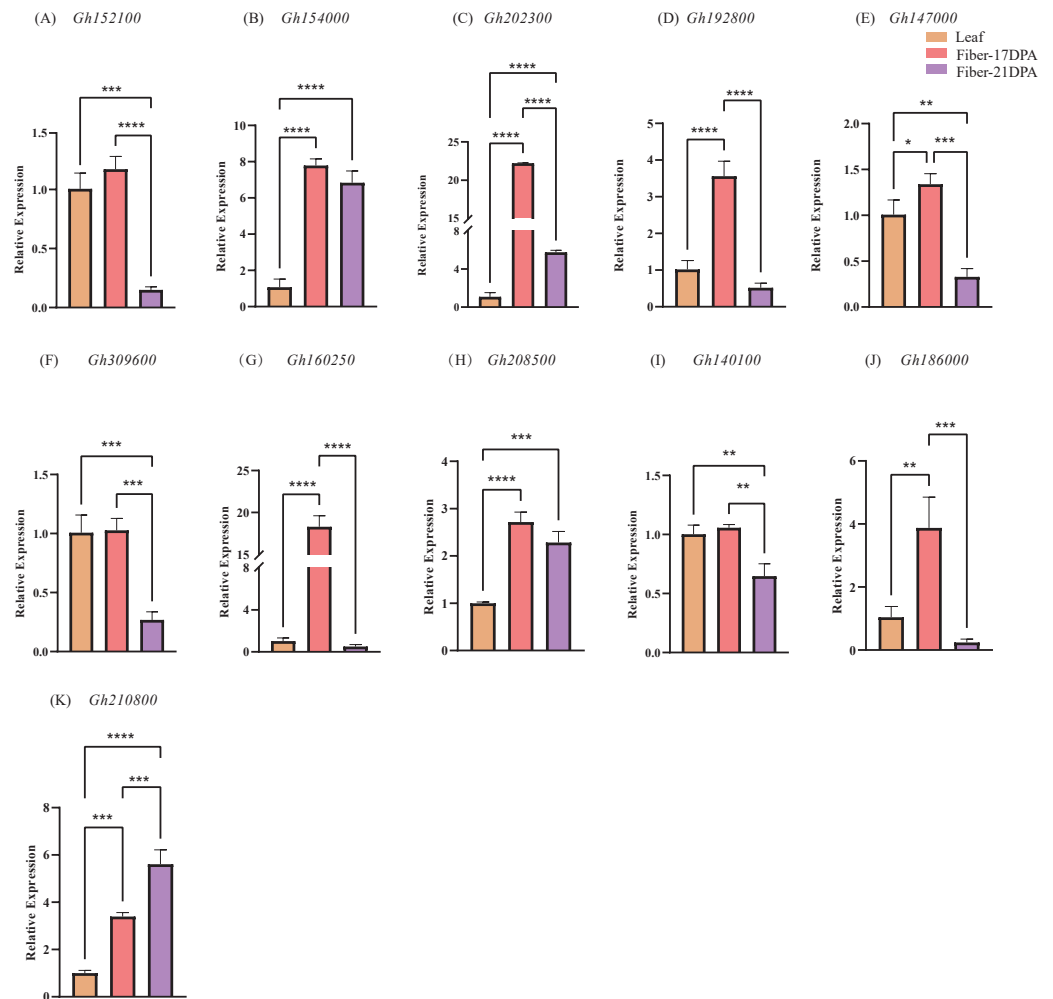
Figure 6 **The relative expression levels of 11 candidate genes in upland cotton.** Note: (A–E) belong to A subgenome, (F–K) belong to D subgenome. The asterisk at the top of the bar indicates the statistically significant difference between the various periods (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$).

Full-size ⤢ DOI: 10.7717/peerj.12733/fig-6

interphase and mitosis of mammalian cells, CLASP1 localizes to the centrosome and to the plus ends of MTs (*Maiato et al., 2003*). Studies have shown that in *Arabidopsis thaliana*, the lack of function of *AtCLASP* inhibits the transport of growth hormone and other substances; therefore, *Arabidopsis* shows phenotypes similar to plants lacking growth hormone, such as short plant type, multiple root branching, slow plant development and short pod. Microtubules are channels for plant material transportation (*Brandizzi & Wasteneys, 2013*; *Kakar et al., 2013*; *Zhang, Zhang & Zhao, 2013*). How CLASPs in cotton regulate the involvement of microtubules in the transportation of intracellular substances remains to be further studied.

Due to differences in CLASP functional domains, their family members in upland cotton may have different functions. Recent cotton genomics and genetics development has enabled us to systematically study the *CLASP_N* genes and explore their potential

functions in the biological clock (*Wang et al., 2020*). In the phylogenetic tree, the 16 upland cotton *CLASP_N* genes were divided into 3 subtypes (Fig. 3). Among them, *Gh160250*, *Gh053900*, and *Gh053074* belong to the third subgroup. The structure of the *Gh160250* gene is very different from that of the other genes. This difference in the evolutionary process may be related to subfunctionalization and natural selection of species. Due to the high collinearity of genes in subgroup A and subgroup D of the tetraploid upland cotton genome, 16 members of the *CLASP_N* genes in upland cotton had multiple pairs of collinearity genes (*Chen et al., 2020*). Homologous genes always have the same biological function in the evolution stage, and their exon-intron structure and motif distribution are the same (*Altenhoff & Dessimoz, 2009*).

An analysis of chromosome location and genome collinearity indicated that the hybridization of upland cotton A and D subgenomes involves gene amplification through tandem duplication and fragment duplication (*Jackson & Chen, 2010*). A high degree of collinearity occurs between the *CLASP_N* genes of At and the Dt subgenome of the tetraploid *G. hirsutum*. In this study, many of the 16 CLASP_N family members were homologous genes, indicating that upland cotton has undergone large-scale gene rearrangements at the genome level during speciation (*Wang et al., 2018*). For the 16 genes identified, 11 genes were extracted for qRT-PCR analysis. Comparison of gene expression in different tissues showed that the gene expression level in leaves is generally lower than that in fiber-17DPA; meanwhile, the expression level in leaves is higher than that of fiber-21DPA for 4 genes, whereas lower for the rest 7 genes. A comparison of the gene expression at fiber-17 DPA and fiber-21 DPA showed that the trend of gene expression level in A subgenome (Figs. 6A–6E) was basically similar to that of D subgenome (Figs. 6F–6J) with a decreasing trend in fiber-21 DPA comparing to fiber-17 DPA except for *Gh210800* (Fig. 6K), indicating that most of these genes played important roles in the early phase of fiber growth.

An analysis of the conserved domains showed that the *Gh053074* gene contained only motif 1, motif 2 and motif 4 while the rest of the genes contained more motifs. Among them, the structures of the two subgroups containing *Gh160200* and *Gh053900* were similar, indicating that they may have similar functions (Fig. 3). Systematic analysis and research can be carried out according to the subgroup classification on the gene expression pattern map. The analysis of the cis-acting elements of the promoters revealed that almost every gene has phytohormone response elements. Phytohormones play an important role in regulating plant growth and development. For example, the ABRE is an abscisic acid response element in plants that binds to transcription factors to promote or inhibit the expression of abscisic acid-induced genes. It has been confirmed in *Arabidopsis* that this element is related to the stress resistance of plants (*Pan et al., 2020*). Other elements related to adversity are also included in this research, *e.g.*, the cis-acting element involved in the MeJA response is CGTCA-Motif, and the cis-acting element involved in the low-temperature response is LTR. TCA-element is a cis-acting element involved in the salicylic acid response, TC-rich repeat is a cis-acting element involved in defense and stress responses, P-box is a gibberellin response element, and TGA-element is an auxin-responsive element. The homeopathic element involved in the MYB transcription factor-binding site under drought stress is MBS.

The CGTCA motif is a cis-acting regulatory element involved in MeJA responsiveness. The regulation of MeJA during flower development involves various physiological processes, such as filament elongation, pistil development, anther development and anther dehiscence. Among a variety of plants, such as wheat, rice, and rape, MeJA plays an important role in regulating anther dehiscence. Therefore, plant hormone response elements in the promoter sequence of the *CLASP_N* genes may regulate the development of anthers in cotton. Elements involved in hormone regulation have been found in many promoter regions, including ABREs, CGTCA motifs, P-boxes, TCA elements, and TGA elements (*Chen & Liu, 2019*). These homeopathic elements likely play a very important role in the development of cotton. In summary, we can infer that *CLASP_N* genes in upland cotton may have a certain regulatory role under hormonal and adversity stress. Upland cotton fiber development can be divided into four stages, and many genes are involved in the development and regulation of fibers at each stage of fiber development. During the secondary thickening of the cotton fiber cell wall, the number of periplasmic microtubules increased significantly, and their arrangement direction was consistent with the direction of the newly formed fiber layer, whereas the synthesis of the cotton fiber cell secondary wall was related to the arrangement pattern of the periplasmic microtubules (*Dixit & Cyr, 2004*). Therefore, these genes may affect the development of cotton fiber by regulating the dynamic reorganization of microtubules and the change in the arrangement direction. According to previous studies, the *CLASP_N* genes can act on microtubules and cell membranes at the same time, which may be related to the transportation of intracellular substances (*Yaffe, Stuurman & Vale, 2003*). Previous work showed that CLASP_N family genes were located on the mitochondrial membrane. CLASPs may regulate microtubules, affect the structure, distribution or function of mitochondria in fiber cells, and participate in the regulation of fiber development. By regulating the orientation and dynamic distribution of microtubules, CLASPs participate in the regulation of fiber development, affects fiber elongation and secondary growth, and finally affects fiber quality.

CONCLUSIONS

Based on genomic data of upland cotton, a genome-wide identification analysis of upland cotton CLASP_N family was performed, and 16 family members were identified. A comprehensive analysis of their phylogenetic relationships, gene structures, evolution processes, promoter sequences, and expression patterns revealed their contribution to cotton fiber quality. These results will facilitate further investigations of the molecular role of *CLASP_N* genes in cotton fiber development.

ADDITIONAL INFORMATION AND DECLARATIONS

the Practice Innovation Training Program Projects for College Students (2021117). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Meijun Ji performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Kangtai Sun, Hui Fang and Zhimin Zhuang performed the experiments, prepared figures and/or tables, and approved the final draft.
- Haodong Chen and Allah Ditta analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Qi Chen, Ziyi Cao and Yiting Wang analyzed the data, prepared figures and/or tables, and approved the final draft.
- Muhammad Kashif Riaz Khan, Kai Wang and Baohua Wang conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12733#supplemental-information.

REFERENCES

Aflitos SA, Schijlen E, Finkers R, Smit S, Wang J, Zhang GY, Li N, Mao LK, Jong HD, Bakker F, Gravendeel B, Breit T, Dirks R, Huits H, Struss D, Wagner R, Leeuwen HV, Ham RV, Fito L, Guigner L, Sevilla M, Ellul P, Ganko EW, Kapur A, Reclus E, Geus BD, Geest HV, Hekkert BT, Haarst JC, Smits L, Koops A, Perez GS, Ridder D, Heusden S, Visser R, Quan ZW, Min JM, Liao L, Wang XL, Wang GB, Yue Z, Yang XH, Xu N, Schranz E, Smets EF, Vos RA, Rauwerda H, Ursem R, Schuit C, Kerns M, Berg J, Vriezen WH, Janssen A, Jahrman A, Moquet F, Bonnet J, Peters SA. 2015. Exploring genetic variation in the tomato (*Solanum section Lycopersicon*)

clade by whole-genome sequencing. *Plant Journal for Cell & Molecular Biology* **80(1)**:136–148 DOI 10.1111/tpj.12616.

Al-Bassam J, Kim H, Brouhard G, Oijen AV, Harrison SC, Chang F. 2010. CLASP promotes microtubule rescue by recruiting tubulin dimers to the microtubule. *Developmental Cell* **19**:245–258 DOI 10.1016/j.devcel.2010.07.016.

Alkuddsi Y, Gururaja-Rao MR, Patil SS, Joshi M, Gowda TH. 2013. Heterosis studies and per se performance of intra *hirsutum* hybrids (*G. hirsutum* × *G. hirsutum*) for kapas yield and its components in cotton. *Plant Gene and Trait* **4**:124–141 DOI 10.5376/cgg.2013.04.0006.

Altenhoff AM, Dessimoz C. 2009. Phylogenetic and functional fssessment of orthologs inference projects and methods. *PLOS Computational Biology* **5**:100026 DOI 10.1371/journal.pcbi.1000262.

Ambrose JC, Shoji T, Kotzer AM, Pighin JA, Wasteneys GO. 2007. The *Arabidopsis* CLASP gene encodes a microtubule-associated protein involved in cell expansion and division. *The Plant Cell* **19**:2763–2775 DOI 10.1105/tpc.107.053777.

Applequist WL, Cronn R, Wendel JF. 2001. Comparative development of fiber in wild and cultivated cotton. *Evolution & Development* **3(1)**:3–17 DOI 10.1046/j.1525-142x.2001.00079.x.

Arce-Rodríguez ML, Martínez O, Ochoa-Alejo N. 2021. Genome-wide identification and analysis of the MYB transcription factor gene family in chili pepper (*Capsicum spp*). *International Journal of Molecular Sciences* **22(5)**:2229 DOI 10.3390/ijms22052229.

Bailey TL, Nadya W, Chris M, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* **34**:369–373 DOI 10.1093/nar/gkl198.

Brandizzi F, Wasteneys GO. 2013. Cytoskeleton-dependent endomembrane organization in plant cells: an emerging role for microtubules. *Plant Journal* **75**:339–349 DOI 10.1111/tpj.12227.

Caroline F, Verena S, Jennifer O, Johannes L. 2014. A TOGL domain specifically targets yeast CLASP to kinetochores to stabilize kinetochore microtubules. *Journal of Cell Biology* **55**:5–571 DOI 10.1083/icb.201310018.

Chen P, Liu QZ. 2019. Genome-wide characterization of the WRKY gene family in cultivated strawberry (*Fragaria* × *ananassa* Duch.) and the importance of several group III members in continuous cropping. *Scientific Reports* **9**:8423 DOI 10.1038/s41598-019-44479-7.

Chen ZJ, Scheffler BE, Dennis E, Triplett BA, Zhang T, Guo W, Chen X, Stelly DM, Rabinowicz PD, Town CD, Arioli T, Brubaker C, Cantrell RG, Lacape JM, Ulloa M, Chee P, Gingle AR, Haigler CH, Percy R, Saha S, Wilkins T, Wright RJ, Van Deynze A, Zhu Y, Yu SI. 2007. Abdurakhmonov, I. Katageri, P.A. Kumar, R. Mehboob Ur, Y.Zafar, J.Z. Yu, R.J. Kohel, J.F. Wendel, A.H. Paterson, Toward sequencing cotton (*Gossypium*) genomes. *Plant Physiology* **145**:1303–1310 DOI 10.1104/pp.107.107672.

Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding MQ, Ye WX, C. Kirkbride R, Jenkins J, Plott C, Lovell J, Lin YM, Robert V, Liu

B, Simpson S, E. Scheffler B, Wen L, A. Saski C, E. Grover C, Hu GJ, L. Conover J, W. Carlson J, Shu SQ, B. Boston L, Williams M, G. Peterson D, McGee K, C. Jones D, F. Wendel J, M. Stelly D, Grimwood J, Schmutz J. 2020. Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. *Nature Genetics* **52**(5):525–533 DOI 10.1038/s41588-020-0614-5.

Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S. 2017. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant Journal* **89**(4):789–804 DOI 10.1111/tpj.13415.

Cleavage Akhmanova A, Hoogenraad CC. 2005. Microtubule plus-end-tracking proteins: mechanisms and functions. *Current Opinion in Plant Biology* **17**:47–54 DOI 10.1016/j.ceb.2004.11.001.

Dixit R, Cyr R. 2004. Encounters between dynamic cortical microtubules promoter ordering of the cortical array through angle dependent modifications of microtubule behavior. *The Plant Cell* **16**:3274–3284 DOI 10.1105/tpc.104.026930.

Eng RC, Schneider R, Matz TW, Carter R, Sampathkumar A. 2021. KATANIN and CLASP function at different spatial scales to mediate microtubule re-sponse to mechanical stress in *Arabidopsis* cotyledons. *Current Biology* **31**:1–13 DOI 10.1016/j.cub.2021.05.019.

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 10.1186/1471-2164-15-838.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-V egas A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**:279–285 DOI 10.1093/nar/gkv1344.

Gardiner J. 2013. The evolution and diversification of plant microtubule-associated proteins. *Plant Journal* **75**(2):219–229 DOI 10.1111/tpj.12189.

Groves MR, Hanlon N, Turowski P, Hemmings BA, Barford D. 1999. The structure of the protein phosphatase 2A 462 PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Canadian Metallurgical Quarterly* **96**:99–110 DOI 10.1016/S0092-8674(00)80963-0.

Hamada T. 2014. Microtubule organization and microtubule-associated proteins in plant cells. *International Review of Cell and Molecular Biology* **312**:1–52 DOI 10.1016/B978-0-12-800178-3.00001-4.

He C, Liu X, Silva J, Wang H, Duan J. 2021. Characterization of LEA genes in Dendro-bium officinale and one gene in induction of callus. *Journal of Plant Physiology* **258–259**:153356 DOI 10.1016/j.jplph.2020.153356.

Hu B, Jin J, Guo AY, He Z, Luo J, Gao G. 2015. GSDS 2.0: an upgraded gene feature visu-alization server. *Bioinformatics* **31**(8):1296–1297 DOI 10.1093/bioinformatics/btu817.

Jackson S, Chen ZJ. 2010. Genomic and expression plasticity of polyploidy. *Current Opinion in Plant Biology* **13**:153–159 DOI 10.1016/j.pbi.2009.11.004.

Kakar K, Zhang HT, Scheres B, Dhonukshe P. 2013. CLASP mediated cortical microtubule organization guides PIN polarization axis. *Nature* **495**:529–533 DOI 10.1038/nature11980.

Kawahara Y, Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu JZ, Zhou SG, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T. 2013. Improvement of the *Oryza sativa Nipponbare* reference genome using next generation sequence and optical map data. *Rice* **6(1)**:4 DOI 10.1186/1939-8433-6-4.

Kelly CM, Osorio-Marin J, Kothari N, Hague S, Dever JK. 2019. Genetic improvement in cotton fiber elongation can impact yarn quality. *Industrial Crops and Products* **129**:1–9 DOI 10.1016/j.indcrop.2018.11.066.

Kim HJ, Triplett BA. 2001. Cotton fiber growth in planta and in vitro. Models for plant cell elongation and cell wall biogenesis. *Plant Physiology* **127(4)**:1361–1366 DOI 10.1104/pp.127.4.1361.

Kong ZS, Ioki M, Braybrook S, Li SD, Ye ZH, Lee YRJ, Hotta T, Chang A, Tian J, Wang GD, Liu B. 2015. Kinesin-4 functions in vesicular transport on cortical microtubules and regulates cell wall mechanics during cell elongation in plants. *Molecular Plant* **8**:1011–1023 DOI 10.1016/j.molp.2015.01.004.

Laryssa H, Katherine G, Geoffrey W. 2020. The microtubule-associated protein CLASP is translationally regulated in light-dependent root apical meristem growth. *Plant Physiology* **184(4)**:4 DOI 10.1104/pp.20.00474.

Le PY, Ambrose C. 2018. CLASP promotes stable tethering of endoplasmic microtubules to the cell cortex to maintain cytoplasmic stability in *Arabidopsis* meristematic cells. *PLOS ONE* **13(6)**:e0198521 DOI 10.1371/journal.pone.0198521.

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Yves VDP, Pieree R, Stephane R. 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.

Letunic I. 2017. SMART. *Available at http://smart.embl-heidelberg.de/* (accessed on 03 March 2017).

Li H, Xu G, Yang C, Yang L, Liang Z. 2019. Genome-wide identification and expression analysis of HKT transcription factor under salt stress in nine plant species. *Ecotoxicology and Environmental Safety* **171**:435–442 DOI 10.1016/j.ecoenv.2019.01.008.

Ma H, Zhao J. 2010. Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (Oryza Sativa L.). *Journal of Experimental Botany* **61**:2647–2668 DOI 10.1093/jxb/erq104.

Maiato H, Fairley EA, Rieder CL, Swedlow JR, Sunkel CE, Earnshaw WC. 2003. Human CLASP1 is an outer kinetochore component that regulates spindle microtubule dynamics. *Cell* **113**:891–904 DOI 10.1016/S0092-8674(03)00465-3.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015.

CDD: ncbi's conserved domain database. *Nucleic Acids Research* **43**:D222–D226 DOI 10.1093/nar/gku1221.

May OL. 1999. Registration of PD94042 germplasm line of Upland cotton with high yield and fiber maturity. *Crop Science* **39**:597–598 DOI 10.2135/cropsci1999.0011183X003900020071x.

Mineo SH, Takahashi NK, Yamada-Hara MK, Tsuzuno T, Aoki-Nonaka YK, Tabeta KC. 2021. Rice bran-derived protein fractions enhance sulforaphane-induced anti-oxidative activity in gingival epithelial cells. *Archives of Oral Biology* **24**:129 DOI 10.1016/j.archoralbio.2021.105215.

Pan WC, Zheng PP, Zhang C, Wang W, Li YP, Fan TT, Liu YS, Cao SQ. 2020. The effect of ABRE BINDING FACTOR 4-mediated FYVE1 on salt stress tolerance in Arabidopsis. *Plant Science* **296**:110489 DOI 10.1016/j.plantsci.2020.110489.

Pietra S, Gustavsson A, Kiefer C, Kalmbach L, Horstedt P, Ikeda Y, Stepanova AN, Alonso JM, Grebe M. 2013. Arabidopsis SABRE and CLASP interact to stabilize cell division plane orientation and planar polarity. *Nature Communications* **4**:2779 DOI 10.1038/ncomms3779.

Sedbrook JC. 2004. MAPs in plant cells: delineating microtubule growth dynamics and organization. *Current Opinion in Plant Biology* **7**:632–640 DOI 10.1016/j.pbi.2004.09.017.

Shi Y, Li W, Li A, Ge R, Zhang B, Li J, Liu G, Li J, Liu A, Shang H, Gong J, Gong W, Yang Z, Tang F, Liu Z, Zhu W, Jiang J, Yu X, Wang T, Wang W, Chen T, Wang K, Zhang Z, Yuan Y. 2015. Constructing a high-density linkage map for *Gossypium hirsutum* × *Gossypium barbadense* and identifying QTLs for lint percentage. *Journal of Integrative Plant Biology* **57**:450–467 DOI 10.1111/jipb.12288.

Song Z, Chen Y, Zhang C, Zhang J, Zhang J. 2019. RNA-seq reveals hormone-regulated synthesis of non-cellulose polysaccharides associated with fiber strength in a single-chromosomal-fragment-substituted upland cotton line. *The Crop Journal* **27**:3–286 DOI 10.1016/j.cj.2019.11.003.

Sudhir K, Glen S, Koichiro T. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**:1870–1874 DOI 10.1093/molbev/msw054.

Sumino Y, Nagai KH, Shitaka Y, Dan T, Yoshikawa K, Chaté H, Oiwa K. 2012. Large-scale vortex lattice emerging from collectively moving microtubules. *Nature* **483(7390)**:448–452 DOI 10.1038/nature10874.

Tang K, Liu JY. 2017. Molecular characterization of *GhPLDα1* and its relationship with secondary cell wall thickening in cotton fibers. *Acta Biochimica et Biophysica Sinica* **49**:33–43 DOI 10.1093/abbs/gmw113.

Voorrips RE. 2002. Mapchart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93(1)**:77–78 DOI 10.1093/jhered/93.1.77.

Wang JJ, Du ZH, Huo XH, Zhou J, Chen Y, Zhang JX, Pan A, Wang XY, Wang F, Zhang J. 2020. Genome-wide analysis of PRR gene family uncovers their roles in circadian rhythmic changes and response to drought stress in *Gossypium hirsutum* L. *PeerJ* **8**:e9936 DOI 10.7717/peerj.9936.

Wang H, Huang C, Zhao W, Dai B, Shen C, Zhang B, Li D, Lin Z. 2016b. Identification of QTL for fiber quality and yield traits using two immortalized backcross populations in upland cotton. *PLOS ONE* **11**:e0166970 DOI 10.1371/journal.pone.0166970.

Wang Y, Tang H, De Barry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40(7)**:e49 DOI 10.1093/nar/gkr1293.

Wang B, Tseng E, Regulski M, Clark TA, Hon T, Jiao Y, Lu Y, Olson A, Stein JC, Ware D. 2016a. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nature Communications* **7**:11708 DOI 10.1038/ncomms11708.

Wang MJ, Tu LL, Yuan DJ, Zhu D, Shen C, Li JY, Liu FY, Pei LL, Wang PC, Zhao GN, Ye ZX, Huang H, Yan FL, Ma YZ, Zhang L, Liu M, You JQ, Yang YC, Liu ZP, Huang F, Li BQ, Qiu P, Zhang QH, Zhu LF, Jin SX, Yang XY, Min L, Li GL, Chen LL, Zheng HK, Lindsey K, Lin ZX, Udall JA, Zhang XL. 2018. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and *Gossypium barbadense*. *Nature Genetics* **51**:224–229 DOI 10.1038/s41588-018-0282.

Yaffe MP, Stuurman N, Vale RD. 2003. Mitochondrial positioning in fission yeast is driven by association with dynamic microtubules and mitotic spindle poles. *Proceedings of the National Academy of Sciences of the United States of America* **100**:11424–11428 DOI 10.1073/pnas.1534703100.

Yin H, You L, Pasqualone D, Kopski KM, Huffaker TC. 2002. Stu1p is physically associated with β-tubulin and is required for structural integrity of the mitotic spindle. *Molecular Biology of the Cell* **13**:1881–1892 DOI 10.1091/mbc.01-09-0458.

Zhang C, Raikhel NV, Hicks GR. 2013. CLASPing microtubules and auxin transport. *Developmental Cell* **24**:569–571 DOI 10.1016/j.devcel.2013.03.008.

Zhang J, Wang JJ, Du ZH, Huo XH, Zhou J, Chen Y, Zhang JX, Pan A, Wang XY, Wang F. 2020. Genome-wide analysis of PRR gene family uncovers their roles in circadian rhythmic changes and response to drought stress in *Gossypium hirsutum L. PeerJ* **8(4)**:14 DOI 10.7717/peerj.9936.

Zhang C, Zhang H, Zhao Y. 2013. Genome-wide analysis of the CCCH zinc finger gene family in Medicago truncatula. *Plant Cell Reports* **32(10)**:1543–1555 DOI 10.1186/1471-2164-9-44.

Zhu SH, Xue F, Li YJ, Liu F, Zhang XY, Zhao LJ, Sun YQ, Zhu QH, Sun J. 2018. Identification and functional characterization of a microtubule-associated protein, ghclasp2, from upland cotton (*Gossypium hirsutum L.*). *Frontiers in Plant Science* **166**:4–462 DOI 10.3389/FPLS.2018.00882X.