

RESEARCH

Open Access



Genome wide identification and characterization of MAPK genes reveals their potential in enhancing drought and salt stress tolerance in *Gossypium hirsutum*

SADAU Salisu Bello^{1,3*} , MEHARI Teame Gereziher¹, AHMAD Adeel¹, TAJO Sani Muhammad¹, IBRAHIM Sani², IQBAL Muhammad Shahid¹, ELASAD Mohammed⁴, ZHANG Jingjing¹, WEI Hengling^{1*} and YU Shuxun¹

Abstract

Background: The cotton crop is universally considered as protein and edible oil source besides the major contributor of natural fiber and is grown in tropical and subtropical regions around the world. Unpredicted environmental stresses are becoming significant threats to sustainable cotton production, ultimately leading to a substantial irreversible economic loss. Mitogen-activated protein kinase (MAPK) is generally considered essential for recognizing environmental stresses through phosphorylating downstream signal pathways and plays a vital role in numerous biological processes.

Results: We have identified 74 MAPK genes across cotton, 41 from *G. hirsutum*, 19 from *G. raimondii*, whereas 14 have been identified from *G. arboreum*. The MAPK gene-proteins have been further studied to determine their physico-chemical characteristics and other essential features. In this perspective, characterization, phylogenetic relationship, chromosomal mapping, gene motif, *cis*-regulatory element, and subcellular localization were carried out. Based on phylogenetic analysis, the MAPK family in cotton is usually categorized as A, B, C, D, and E clade. According to the results of the phylogenetic relationship, cotton has more MAPKs genes in Clade A than Clade B. The *cis*-elements identified were classified into five groups (hormone responsiveness, light responsiveness, stress responsiveness, cellular development, and binding site). The prevalence of such elements across the promoter region of these genes signifies their role in the growth and development of plants. Seven *GHMAPK* genes (*GH_A07G1527*, *GH_D02G1138*, *GH_D03G0121*, *GH_D03G1517*, *GH_D05G1003*, *GH_D11G0040*, and *GH_D12G2528*) were selected, and specific tissue expression and profiling were performed across drought and salt stresses. Results expressed that six genes were upregulated under drought treatment except for *GH_D11G0040* which is downregulated. Whereas all the seven genes have been upregulated at various hours of salt stress treatment.

Conclusions: RNA sequence and qPCR results showed that genes as differentially expressed across both vegetative and reproductive plant parts. Similarly, the qPCR analysis showed that six genes had been upregulated substantially through drought treatment while all the seven genes were upregulated across salt treatments. The results of this study showed that cotton *GHMPK3* genes play an important role in improving cotton resistance to drought and salt

*Correspondence: sbsadau.ste@buk.edu.ng; weihengling@caas.cn

¹ State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China
Full list of author information is available at the end of the article



© 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/>.

stresses. MAPKs are thought to play a significant regulatory function in plants' responses to abiotic stresses according to various studies. MAPKs' involvement in abiotic stress signaling and innovation is a key goal for crop species research, especially in crop breeding.

Keywords: Cotton, MAPK gene, Drought stress, Salt stress

Background

Cotton (*Gossypium spp.*) has become more important for plant research on polyploidization, phylogeny, cytogenetics, and genomics. It is regarded as one of the most important natural plants with the highest diversity, as well as having the greatest commercial value (Wang et al. 2018). Cotton is mainly cultivated as a source of fiber, food, and feed. *Gossypium hirsutum* L., the tetraploid one, is the largest species among the over 50 in *Gossypium* (Wang et al. 2018). *G. hirsutum* is believed to originate from genetic hybridization amongst an A-genome species that may be sourced from *G. herbaceum* (A1) African origin or might be from Asian cotton, also known as *G. arboreum* (A2) with a D-genome species might be from American origin *G. raimondii*. This tetraploid cotton accounts for around 90% share of worldwide cotton production annually (Page et al. 2013).

Several biotic and abiotic factors such as drought, heat, water logging, salinity, and diseases significantly affect cotton productivity, causing significant losses in the focused agricultural sector productivity. The cotton plant undergoes stress due to its growth habit; that is, it grows according to environmental conditions which are highly affected by environmental stress. Traditional crop breeding techniques have limitations such as crossing barriers, long-time effects, and genetic disease transformation. These lead to the development of recent breeding techniques (Zhang et al. 2014a).

The high amount of greenhouse gas emission in the atmosphere, as well as associated air pollution, is a major cause of heatwaves, floods, and drought. Drought stress can significantly impact crop production, and the severity and duration of the stress are also important factors. The availability of water is a critical factor in achieving long-term sustainability in crop production (Khan et al. 2018). Drought stress is among the factors that affect cotton production, because 50% global cotton supply comes from drought stress-challenged areas of the world. Cotton crops require improved yields and yield balance in both standard and moisture-stressed environments (Tuteja 2017). Drought stress influences cotton plants' growth and productivity by inducing several morphophysiological and biochemical changes. Physiological and metabolic features such as photosynthesis, stomatal conductance, respiration, energy output, carbohydrate metabolism, and ultimately yield are clogged even though

cotton has various mechanisms to relieve and withstand water-deficit stress (Tian et al. 2019). Previous studies showed that abiotic stress treatments such as salt, drought, and low temperature triggered the expression of MAPK2, MAPK3, MKK3, and MAPK16. *GhMAPK16* is thought to be active in the signaling pathways and an Arabidopsis drought tolerance was significantly improved by overexpression (Shi et al. 2011). *GhMKK3* functions in drought tolerance by regulating stomatal responses and root growth (Wang et al. 2016), *GhMAPK16* is involved in disease resistance and drought sensitivity (Shi et al. 2011), *GhMAPK3* enhances cold, drought, and salt stress in *Arabidopsis* (Sadau et al. 2021), and *GhMAPK2* positively regulates salt and drought tolerance in tobacco (Zhang et al. 2011).

Salt stress is one of the most severe environmental stresses that plants face. Roots are the first and most direct organs to detect a signal. From germination to boll formation, salt stress harms cotton physiology, and the tolerance mechanism is well described (Munns and Tester 2008). However, early salinity tolerance responses of plant growth may not be a substantial measure of tolerance for salts across various plant species. Screening for salt tolerance among different plants may use physiological parameters as stress tolerance indicators, while enzyme concentration could be used as salt tolerance assessment in cotton. A few novel approaches, such as transcriptome profile, methylation-sensitive amplified polymorphism analysis at genes and cell levels, and genetic diversity assessed by various molecular markers, revealed salt stress-induced epigenetic changes in cotton cultivars and their salt tolerance mechanism. *GhMAPK6a* negatively regulates osmotic tolerance and bacterial infection in transgenic *Nicotiana benthamiana* and plays a pivotal role in its development (Li et al. 2013). *GhMAPK4* confers the transgenic Arabidopsis hypersensitivity to salt and osmotic stresses (Wang et al. 2015). *GhMKK5* affects disease resistance, induces HR-like cell death, and reduces the tolerance to salt and drought stress in transgenic *Nicotiana benthamiana* (Zhang et al. 2012a; b). *GhMAPK4* negatively regulates salt stress (Wang et al. 2015), while *GhMAPK17* positively regulates salt stress (Zhang et al. 2014a, b).

Plants' adaptive responses to environmental changes triggered by external and internal influences primarily depend on their interpretation of external

signals. Multiple signal transduction pathways are used to amplify these perceived signals. MAPK can be considered as the standard signal regulation mechanism, that transforms external stimuli into cells. Following three sequentially active kinases from MAPK cascades are MAPK (MAP kinase), MAPKK (MAP kinase kinase), and MAPKKK (MAP kinase kinase kinase) (Bengough et al. 2011; Nakagami et al. 2005; Sadau et al. 2021; Wang et al. 2018). MAPK is assumed to significantly identify tolerance for environmental stress across eukaryotes by activating multiple cellular protein receptors (Larade and Storey 2006; Rohila and Yang 2007). MAPKs are activated when plants are stressed (Li et al. 2014; Wang et al. 2016; Zhang et al. 2014a, b). However, most MAPK researches have been done on model plants like *Arabidopsis* and tobacco (Li et al. 2013; Ichimura et al. 2000; Zhang et al. 2013). Also, almost all MAPKs examined so far enhanced abiotic or biotic stress. The importance of MAPK found within the same group differs from *Arabidopsis thaliana AtMPK3 and AtMPK6* (Ichimura et al. 2000). A total of 74 MAPK genes with 41 across *G. hirsutum*, followed by 19 genes in *G. raimondii*, whereas 14 genes present in *G. arboreum* have been identified in this research. The MAPK proteins were studied further to determine their physicochemical characteristics, phylogenetic relationship, chromosomal mapping, and conserved motifs. Further experimental analysis of seven selected *GHMPK* genes was carried out for tissue-specific tissue expression, drought, and salt stresses to confirm their functions in cotton. The findings would give substantial evidence, laying the groundwork for further research into the molecular and biological functions of MAPK in cotton.

Materials and methods

Plant materials, growth conditions, and stress treatments

An upland cotton (TM-1, *G. hirsutum* L.), was used to evaluate tissue/organ and stress expression. Plants were planted under 25 °C, with a light cycle of 16 h light and 8 h dark cycle in a controlled chamber. Tissues like cotyledons, leaves, stems, roots, and fibers (5 days post anthesis, DPA) were collected. A drought stress treatment induced by 15% polyethylene glycol (PEG-6000) was applied to seedlings, and for salt stress, a 200 mmol·L⁻¹ sodium chloride (NaCl) treatment was used (Li et al. 2013; Ma et al. 2020). Sample collection was carried out at 0, 2, 6, and 12 h after treatment. Sampling for each treatment was carried out three times, and samples were immediately stored in liquid nitrogen and preserved under a temperature of -80 °C.

Identification and analysis of physicochemical properties of MAPK genes

Proteins encoded by *MAPK* in *G. hirsutum*, *G. arboreum*, and *G. raimondii* have been obtained from cotton database CottonFGD using the PF00069 obtained from the Pfam database (<http://www.cottonfgd.org/>), considering a significant level as E-value < 0.01. Further confirmation of domains was carried out using the following online tools: <http://prosite.expasy.org/scanprosite/> and <http://smart.embl-heidelberg.de/>. The physicochemical characteristics of the cotton MAPK gene were retrieved from cottonFGD. For localization of *GHMAPK*, *GaMAPK*, and *GrMAPK* protein, the protein sequence was downloaded from cottonFGD, and the prediction was carried out through Wolf PSORT (<https://wolfsort.hgc.jp>) online server.

Sequence alignments, phylogenetic tree construction and collinearity

The MAPK protein sequences obtained from *G. hirsutum*, *G. arboreum*, and *G. raimondii*, and the protein sequence of *A. thaliana* obtained from Phytozome (<https://phytozome.jgi.doe.gov/>) were subjected to the Neighbor-Joining (NJ) approach to investigate the evolutionary relationship. Computer software package MEGA 7.0 was used to construct a phylogenetic tree, considering Jones-Taylor-Thornton to be the substitution model through selecting 1 000 replications. For gene collinearity, *G. hirsutum* protein sequence was considered for blast search across *G. raimondii* and *G. arboreum* protein database considering E-value as < 0.01, and significance were considered ≥ 90 significant. The Gene IDs, GFF3 files, and linked files were used to construct the collinearity using the TBtools software (Chen et al. 2018).

Motif identification, gene structural analysis, chromosomal mapping, and promoter analysis of the cotton MAPK genes

The MEME, an online tool, was used to determine the cotton conserved motifs of MAPK. TB tools software was then used for the motif visualization. The coding sequences (CDS) were compared with the MAPK gene's genome sequences through an online gene structure tool (<http://gsds.gao-lab.org/>). Information about the chromosome was done by extracting cotton GFF3 from cotton GDP (<http://www.cottonfgd.org/>) and then mapped with the gene ID using the TBtools software (Tamura et al. 2011). To examine the role of the *GHMAPK* gene's regulatory region in cotton, an upstream sequence within a 2 000 bp distance from the start codon has been considered and searched for

PlantCARE program (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) (Lescot et al. 2002).

RNA isolation, cDNA synthesis, and qRT-PCR

The RNA isolation was carried out from samples following the kit protocol using an RNA extraction kit (Polysaccharides and Polyphenolics-rich) (Tiangen, China). Using a PrimeScript™ RT reagent package with a gDNA Eraser, RNA has been further transcribed to cDNA. RT-qPCR was performed using actin as a reference gene. The relative expression was calculated using the $2^{-\Delta\Delta C_T}$ method (Schmittgen and Livak 2008). Each experiment has been repeated 3 times, along with three technical replicates.

Expression patterns of *GHMPK* gene in different tissues, under drought, salt, and validation RNA sequencing data

RNA sequence data for TM-1 was obtained from our lab. We analyzed the RNA sequence data under tissue expression, drought, and salt stress. Samples for tissue expression were taken from cotyledon, leaf, stem, root, and fiber at 5 DPA, while the sample for drought and salt were taken at 0, 2, 6, and 12 h as experimental conditions. PEG-6000 has been used for drought induction, whereas sodium chloride (NaCl) solution is used for salt treatment. Log transformation was carried out for reading/kilobase/million mapped values, and heatmap was constructed using software package TB tools.

Results

Identification and physicochemical traits of MAPK gene family

A total of 74 MAPK genes were detected in cotton; out of them, 41 were in *G. hirsutum*, 19 in *G. raimondii*, whereas 14 were in *G. arboreum*. These genes' proteins were studied further to evaluate their physicochemical characteristics as well as other features. These genes were reported to express proteins ranging in length between 87 and 628 amino acids, with molecular mass between 9.988 and 71.503 KDa and isoelectric points (pI) between 5.672 and 10.23. All identified MAPK genes have grand average hydropathy > 0, implying that all MAPK genes in cotton were hydrophilic. Based on the WOLF PSORT (<https://wolfpsort.hgc.jp>) analysis result, GHMAPK proteins are localized across different parts of the cell including the nucleus, cytoplasm, mitochondria, and chloroplast. The proteins were predominantly located on the cytoplasm (Additional file 1: Table S1).

Phylogenetic analysis of MAPK proteins

A phylogenetic tree was constructed based on multiple sequence alignment of 41 GHMAPK, 14 GaMAPK protein sequences, 19 GrMAPK protein sequences, and 18

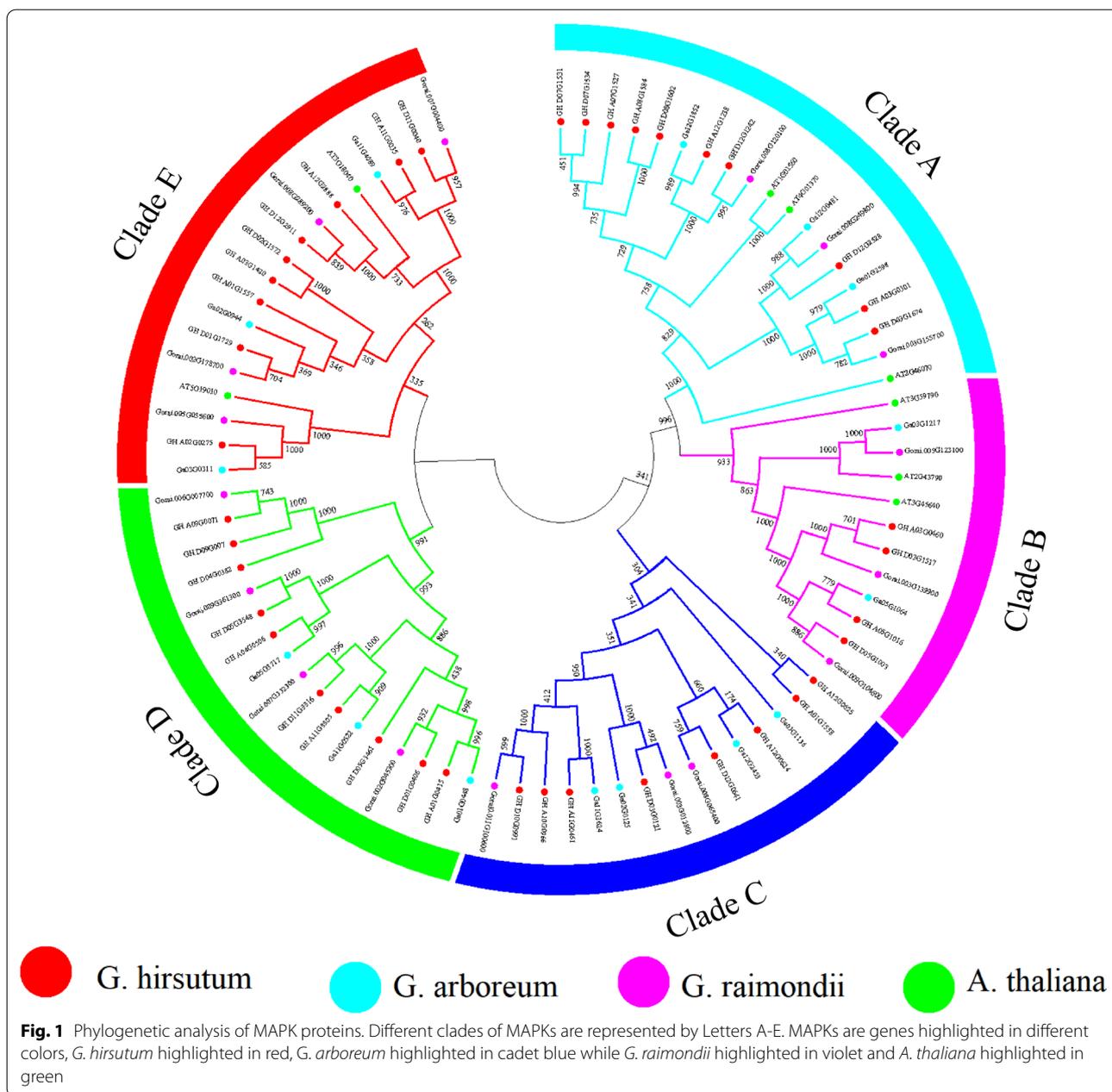
Arabidopsis MAPK protein sequences revealing the evolutionary relationship of cotton MAPKs. The cotton MAPK family is categorized into A, B, C, D, and E clades based on results from the phylogenetic tree (Fig. 1). Clade A comprises of all the genes with TDY phosphorylation site, clade B members contain TEY motif, Clade C members contain TDY motif, clade D members contain more TDY motif and few TEY motifs, and Clade E members contain TEY motif. There are 26 GHMAPKs, 9 GrMAPKS, 9 GaMAPKs, and 3 AtMAPK that contain the TDY motif, while 15 GhMAPK, 8 GrMAPK, 6 GaMAPK, and 5 AtMAPK contain the TEY motifs. These show that the cotton MAPK genes have more TDY motifs and minor TEY motifs while the AtMPK genes have more TEY motifs and minor TDY motifs. The cotton MAPK classification was consistent with previous findings (Zhang et al. 2014b). This indicates that TEY MAPKs motif may significantly function in dicot plants than TDY MAPKs motif (Fig. 1).

Collinearity analysis was performed between the physical maps of the subgenomes about *G. hirsutum*, *G. raimondii*, and *G. arboreum* for associations among A vs. D, A vs. At, and D Vs. Dt subgenome (Fig. 2). Generally, most of the MAPK genes from tetraploid cotton represented high similarities to the D genome (*G. raimondii*) and the A genome (*G. arboreum*).

Motif identification and gene structure analysis of MAPK

Investigations for conserved motif about MAPK proteins were carried out through MEME 41 GHMAPKS, GaMAPKS, and 19 GrMAPKS putative protein sequence was submitted to search for conserved motifs (Fig. 3). As shown in the figure, GHMAPKS, GrMAPKS, and GaMAPKS both possess 20 motifs. In GHMAPKS, the majority of the identified genes contains 10 similar motif composition (1, 2, 3, 4, 5, 6, 10, 13, 17, and 18), few among them contains 9 motifs (7, 8, 9, 11, 12, 14, 15, 16, and 19) and only 3 genes; *GH_A11G0035*, *GH_A12G2888*, *GH_D11G0040*, and *GH_D12G2911* contains motif 20. While in *GaMAPK*, the majority of identified gene contains 9 similar motif composition (1, 2, 3, 4, 6, 8, 9, 10, and 15), few among them contains 5 motifs (5, 7, 16, 17, and 18) only 3 among the genes; *Ga01G0448*, *Ga05G3624*, *Ga11G0522* contain motif 19, and three among the genes; *Ga01G2598*, *Ga12G0481*, and *Ga03G1217* contains motif 20. Lastly, *GrMPK* genes contains 9 similar motif composition (1, 2, 3, 4, 5, 6, 8, 9, 15), few among them contain 10 motifs (7, 10, 11, 12, 13, 16, 17, 18, 19, and 20) and only few among the genes; *Gorai.008G065400*, *Gorai.011G100600* and *Gorai.003G012800* contain motif 14.

Gene structure was examined by using an online tool (<http://gsds.gao-lab.org/>) for cotton MAPKs. In *G. hirsutum*, out of 41 genes, 38 were found to possess an intron,

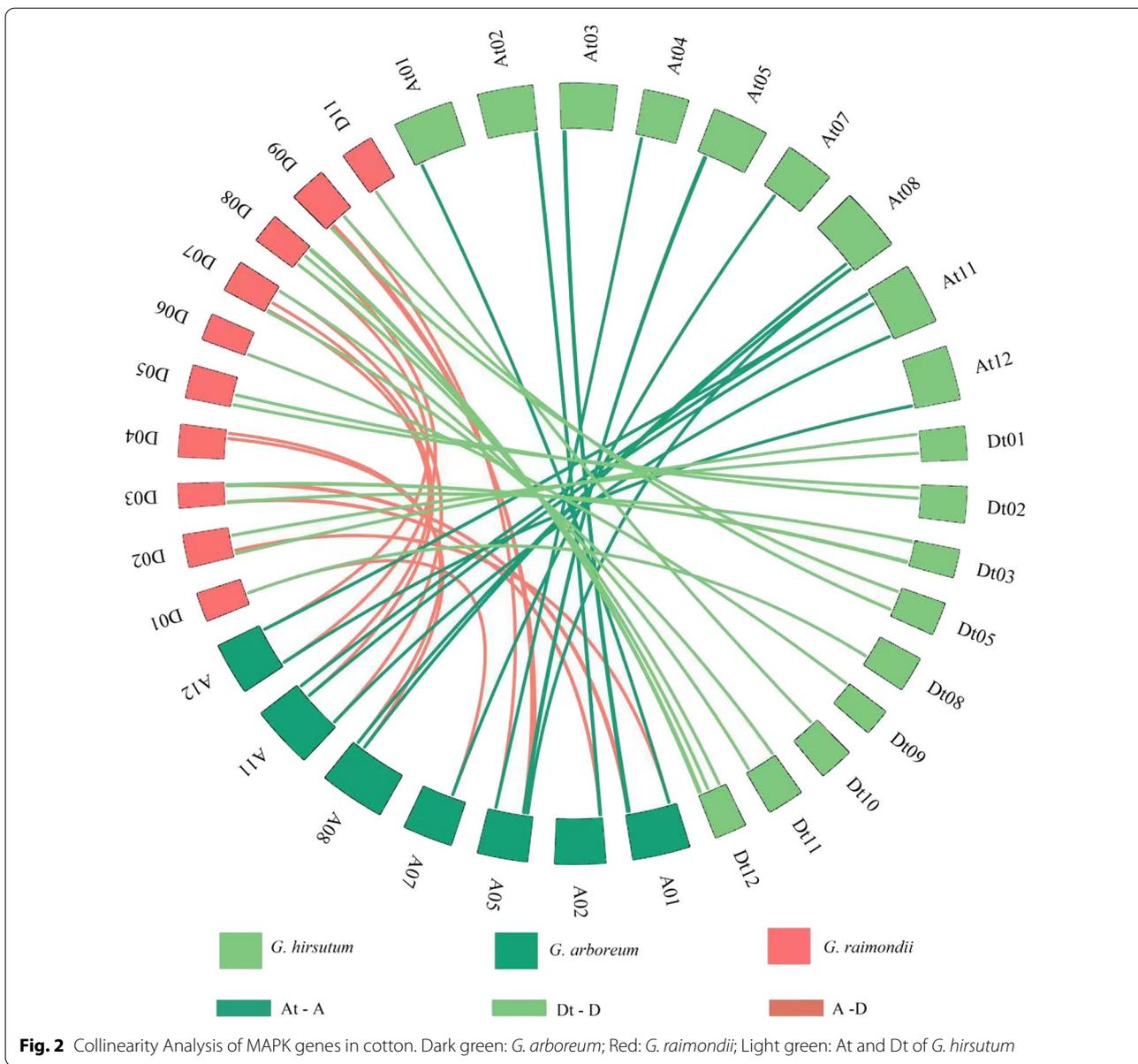


and 3 are intronless. The most extended intron length was observed in *GH_A07G1527* (Fig. 4). In *G. arboreum*, all the 19 genes possess an intron, and the highest intron length was observed in *Ga02G0944* (Fig. 4B). In *G. raimondii*, both genes possess an intron, and the highest intron length is found in *Gorai.006G007700* (Fig. 4C).

Chromosomal mapping of MAPK genes in *Gossypium* species

The chromosome distribution was investigated in the 3 *Gossypium* species and found that among the 41 *G.*

hirsutum MAPK genes, 20 were located in the At-sub genome, and 21 were located in the Dt subgenome. Chromosome A₁₂ and its homologous D₁₂ had the highest gene locus. Three genes were found in A₀₁, A₀₃, A₁₁, D₀₃, and D₀₅, D₀₇, D₀₁ and D₀₂ have two genes each, the remaining A₀₂, A₀₅, A₀₇, A₀₈, A₀₉, A₁₀, D₀₄, D₀₈, D₀₉, and D₁₀ have one gene, respectively (Fig. 5A). In *G. arboreum*, the *GrMPK* genes were mapped on chromosome A₀₁, A₀₂, A₀₃, A₀₅, A₁₁, and A₁₂. The highest gene locus was observed in chromosomes A₀₃, A₁₁, and A₁₂, each with three genes (Fig. 5B). While chromosomes A₀₁, A₀₂, and

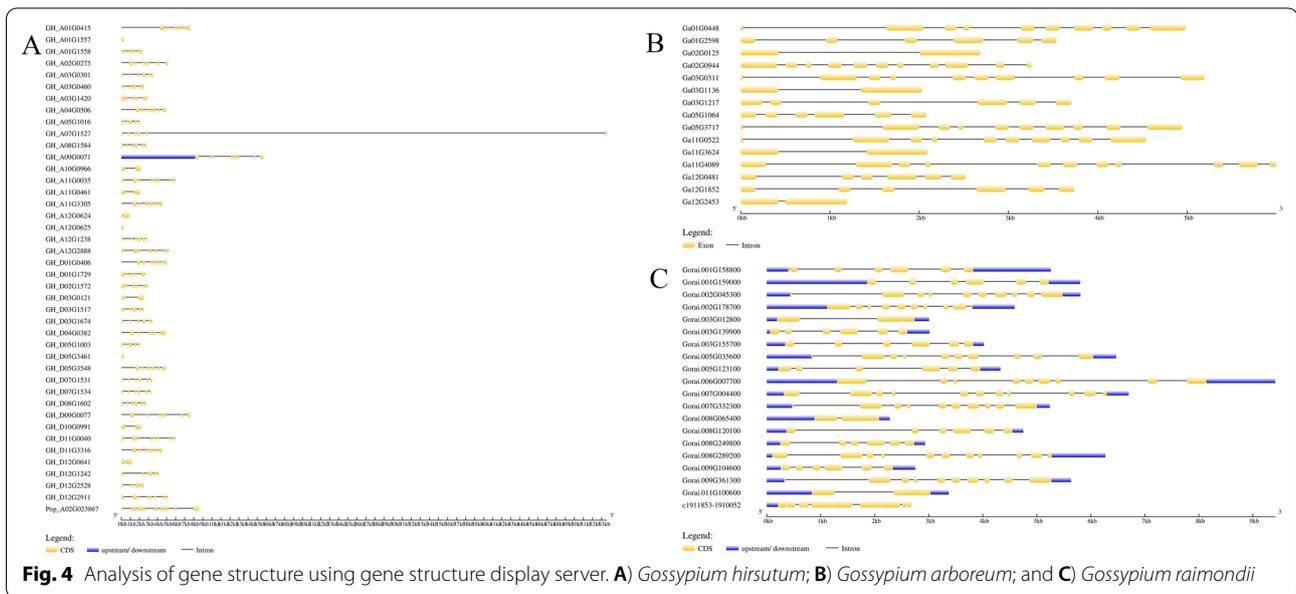
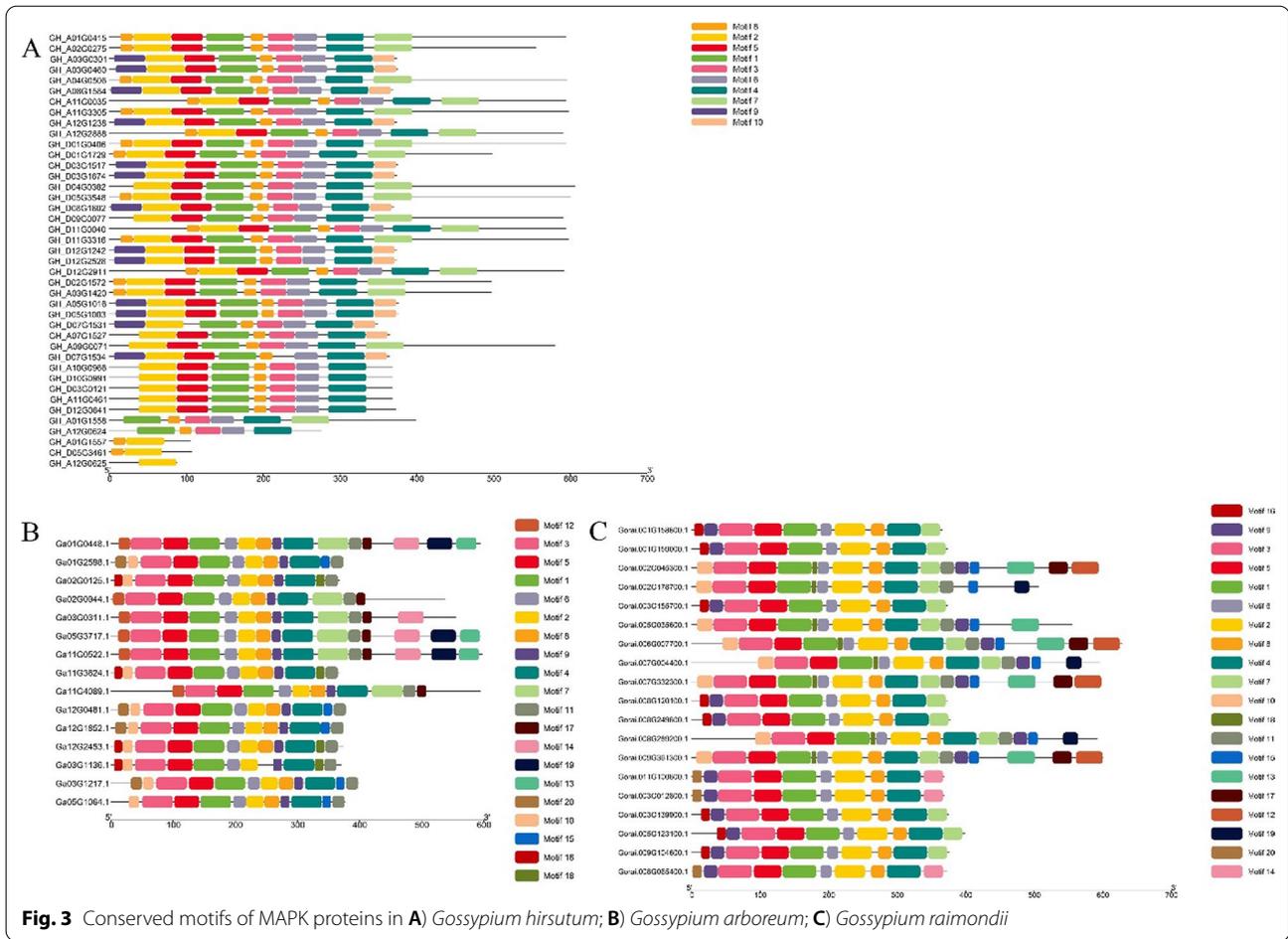


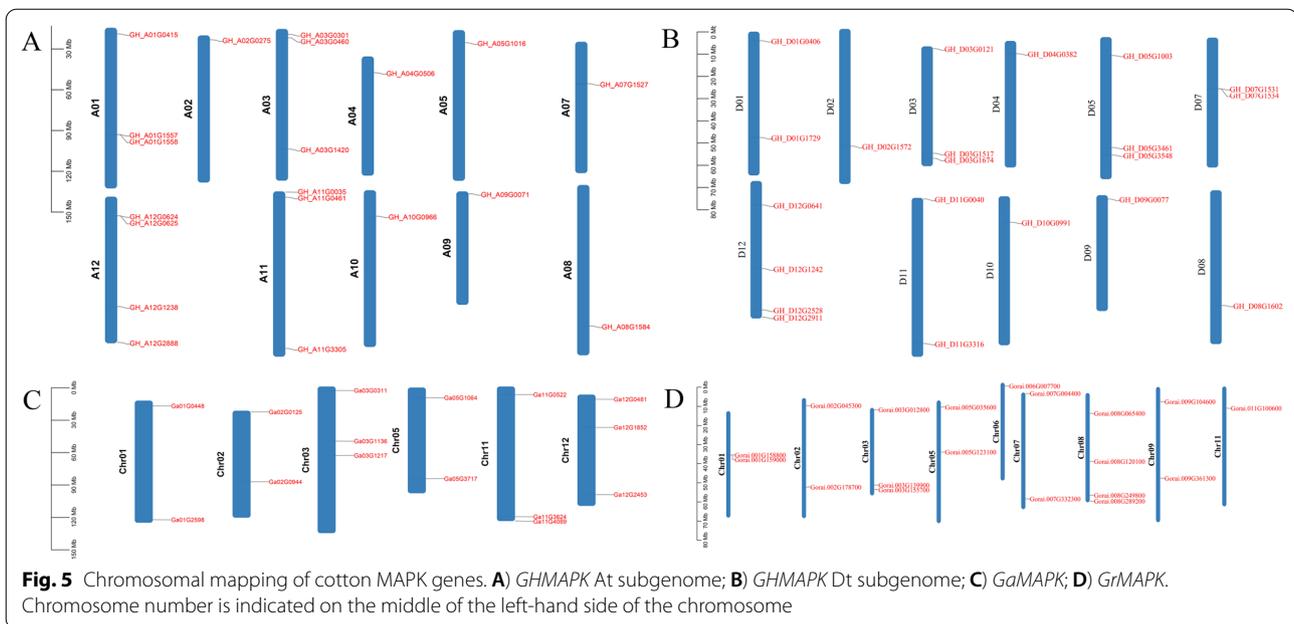
A₀₅ possess two genes each. The genes in *G. raimondii* are distributed among D₀₁, D₀₂, D₀₃, D₀₅, D₀₆, D₀₇, D₀₈, D₀₉, and D₁₁ chromosomes. Chromosome D₀₈ has the most gene loci, followed by D₀₃, three genes (Fig. 5C). While Chromosomes D₀₁, D₀₂, D₀₅, D₀₉ have two chromosomes each, one chromosome was observed in D₀₆ and D₁₁ (Fig. 5C).

Determination of cis-regulatory elements

Cis-regulatory elements are assumed to perform various functions based on their arrangement and location across promoters (Biłas 2016). We have analyzed cotton MAPKs promoter regions for the determination of their

cis-elements. The 2 kb sequences from the start of transcription from each GHMKPS on the upstream side were considered and used. The cis-elements identified in the GHMAPK promoter region were categorized into five functions: hormone responsiveness, stress responsiveness, light responsiveness, cellular responsiveness, and binding site (Additional file 2: Table S2). The majority of GHMAPK genes represented ABRE for the responsiveness of hormone, i.e. cis-element, are involved for elements related to the ABA responsiveness, TGA element (auxin-responsive element), ethylene responsive element (ERE), CA element (salicylic acid-responsive gene element), and GATA-Motif (cis-acting regulatory element





involved in the MeJA-responsiveness). Few among the *GHMAPK* promoters have the GARE motif and P-BOX (gibberellin-responsive). Elements associated with stress include TC-rich repeats (defense and stress-responsive element), LTR (low temperature-responsive element), WUN-motif (wound responsive element), and TC-rich repeats (defense and stress responsiveness element). For light-responsiveness, elements such as ATCT-motif, Box 4, GT1-motif, LAMP-element, TCT-motif, TCCC-motif, CHS-CMA1a, TCT-motif, and TGACG-motif are found. For cellular development and binding site, few *cis*-elements are involved. Several *cis*-regulatory elements related to enhancing tolerance to abiotic and biotic stresses in plants have been found in the promoter sequences in the coding sequence related to the *GHMAPK* gene, suggesting that this gene could be used as an abiotic tolerant gene in cotton (Fig. 6).

RNA sequence analysis and RT-qPCR confirmation of *GHMAPK* under drought and salt treatments

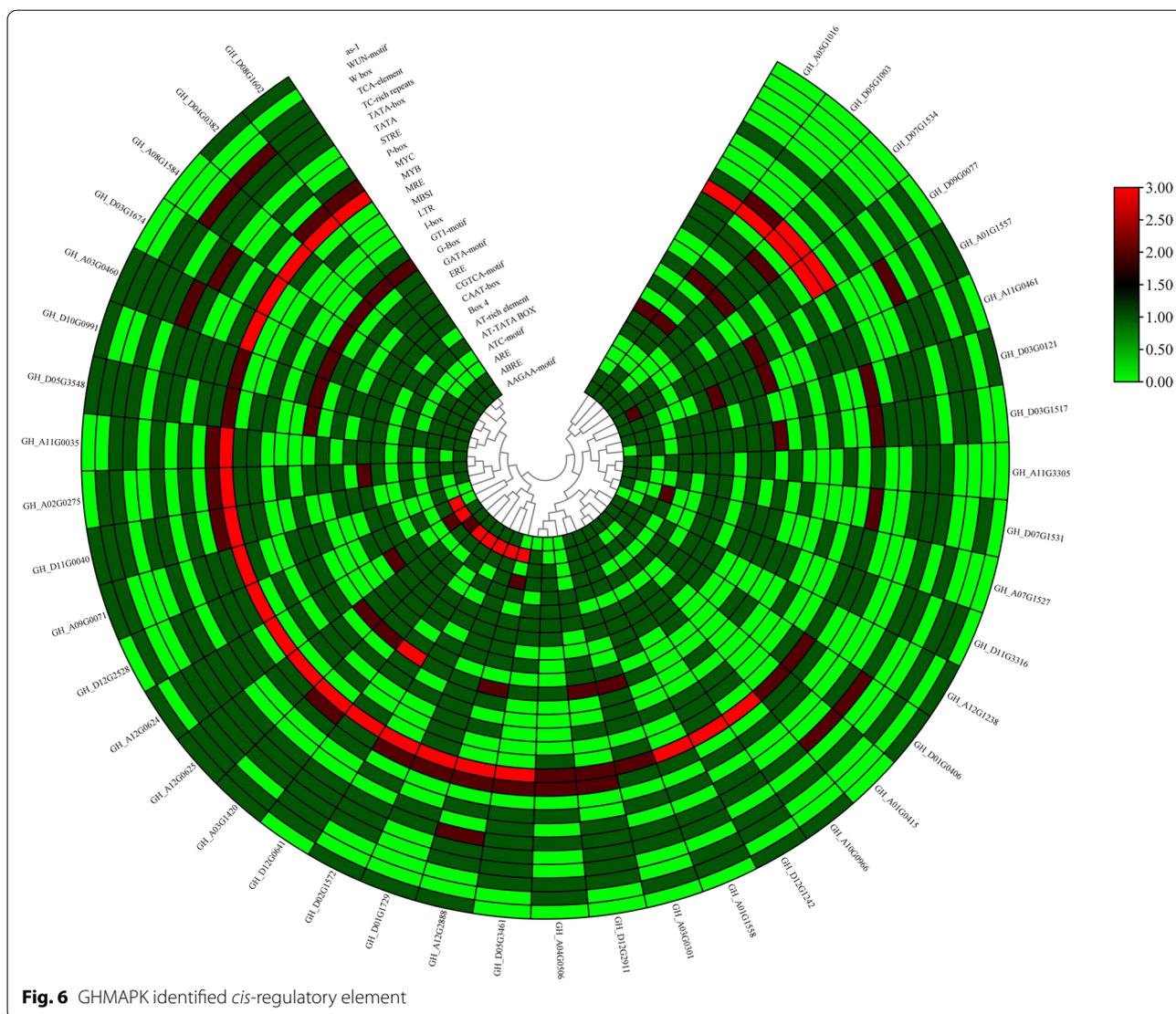
The RNA profile data of TM-1 was used, and the raw data and their transformed log 10 values of the genes were analyzed, and a heat map was constructed. Seven *GHMAPK* genes were found to have different expression patterns across different tissues, including young leaf, true leaf, the cotyledon, stem, fiber, and roots. Gene-specific RT-qPCR primers were designed (Additional file 3: Table S3). Based on RNA-sequence data as well as RT-qPCR results, we determined that *GH_A07G1527* and *GH_D02G1138* are upregulated in all tissues. *GH_D03G0121* shows upregulation in cotyledon, stem, and

root, *GH_D03G1517* and *GH_D05G1003* are upregulated in all the tissues. *GH_D11G0040* gene is upregulated in stem and root, while *GH_D12G2528* shows a relatively low expression (Fig. 7A).

To determine the roles of *GHMAPK* genes under salt and drought stresses, the seven genes have been analyzed for both the RNA-sequence data and RT-qPCR results to detect their expression pattern after treatments. For drought treatment, *GH_A07G1527* was upregulated at 6 h and 12 h, *GH_D02G1138* was upregulated at 2 h, 6 h, and 12 h, and *GH_D03G0121* was upregulated at 12 h. At the same time, *GH_D03G1517* and *GH_D05G1003* were upregulated at 2 h, 6 h, and 12 h. While *GH_D11G0040* showed no expression. Lastly, *GH_D12G2528* was upregulated at 12 h (Fig. 7B). For salt treatment, *GH_A07G1527* and *GH_D02G1138* were upregulated at 2 h, 6 h, and 12 h, *GH_D03G0121* show no expression in RNA seq data and upregulated expression at 12 h post-treatment. *GH_D03G1517* was upregulated at 2 h, 6 h, and 12 h of post-treatment, respectively. *GH_D11G0040* showed low expression in RNA seq data and upregulated at 6 h and 12 h of RT-qPCR. Lastly, *GH_D12G2528* was upregulated at 2 h, 6 h, and 12 h, respectively (Fig. 7C). The results demonstrated that RNA sequence analysis and RT-qPCR expression results represented a strong correlation, with $R^2=0.91$ in drought, $R^2=0.75$ in salt, and $R^2=0.66$ in tissues.

Discussion

Plants are constantly vulnerable to different environmental stress conditions such as pathogen infections, salt, cold, drought, and even oxidative stress. Such



environmental stresses have adverse effects on plant development and productiveness, resulting in significant loss of crop productivity (Tuteja 2017). Unlike animals, plants cannot avoid environmental stress when dealing with complex environmental challenges because plants are sessile and cannot move. To respond to various stress conditions, plants must develop sophisticated pathways that may help them resolve biotic or abiotic signals through appropriate cell signaling mechanisms (Wang et al. 2016). Mitogen-activated protein kinase (MAPK) cascades play an important role in abiotic stress responses as part of a critical signaling transduction module. The MAPK cascade has been identified in plant stress responses and signals transduction in cotton in several previous studies (Wang et al. 2015; Long et al. 2014; Sadau et al. 2021; Shi et al. 2011; Zhang et al. 2014a,

b; Zhang et al. 2020; Zhang et al. 2011). But more studies are needed to strengthen knowledge about the sophisticated biological functions MAPK cascades in cotton under drought and salt stress.

About 41 MAPK genes in *G. hirsutum*, 15 in *G. arboreum*, and 19 in *G. raimondii* have been identified in this study. These three cotton species depict almost identical physicochemical properties about molecular weights with a range of 71.503 to 9.988 KDa and GRAVY values less than zero. Shallow gravity is a strong indication that the protein is hydrophilic and possesses high gravity index. Proteins encoded by stress-responsive genes are closely correlated with hydrophilic. Also, previous researches have confirmed that hydrophilic proteins are involved in tolerating numerous abiotic stresses (Hanin et al. 2011; Mogwomgwa et al. 2019; Lu

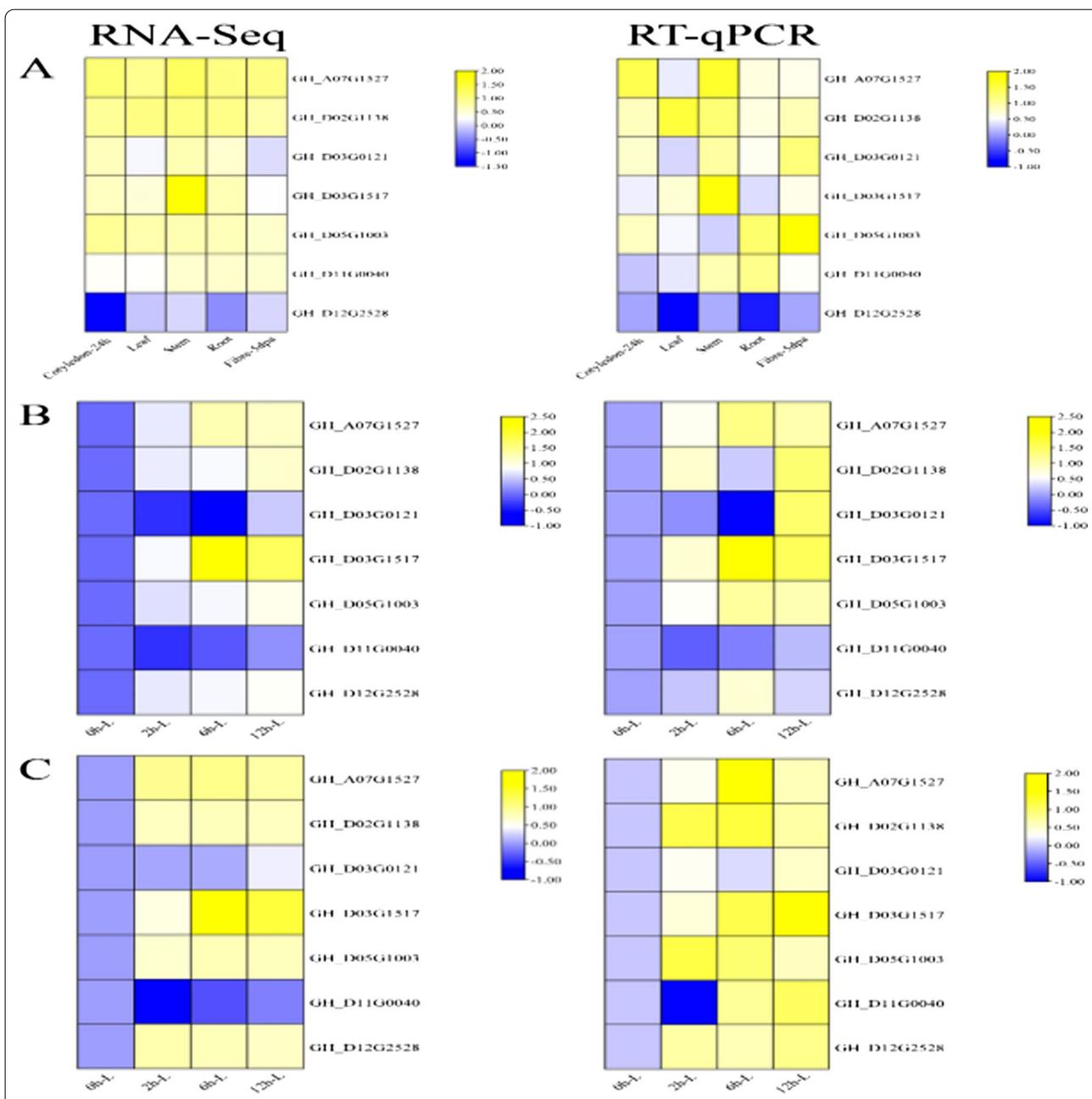


Fig. 7 RNA sequencing analysis and qRT-PCR analysis of the 7 cotton *GHMAPK* genes. **A)** gene expression pattern in cotton tissues; **B)** drought stress gene expression pattern; **C)** salt stress gene expression pattern. The relative expression was determined using the CT method with actin1 serving as a housekeeping gene. The data represents the mean average of three replicates. The TB Tools software was used to build the heat map (shown by log 10 values). Yellow indicates upregulation, blue indicates downregulation, and white indicates no expression

et al. 2019). Cotton has more MAPKS genes in Clade A than Clade B, according to the results of the phylogenetic relationship, which is following the previous research reports in *G. raimondii*, *Arabidopsis*, rice, and poplar (Hamel et al. 2006; Ichimura et al. 2000).

To better understand the potential functions of *GHMAPK* in cotton under various environmental

stresses, we examined the *cis*-element distribution in promoter regions. Based on their position, form, and orientation on the promoter, the *cis*-regulatory elements observed serve various functions. In this study, the *cis*-element identified were classified into five groups (hormone responsiveness, light responsiveness, stress responsiveness, cellular development, and binding site).

The prevalence of such elements across the promoter region of these genes signifies their role in the growth and development of plants (Das and Roychoudhury 2014; Elasad et al. 2018; Escobar-Sepúlveda et al. 2017).

Analysis of exon/intron structure revealed that most of the cotton MAPK genes were interrupted by an intron while few were intronless. The longest intron in *G. hirsutum* is found in *GH_A07G1527*. This results in conformity with previous researches (Zhang et al. 2014a, b; Zhang et al. 2020) which shows most cotton MAPK and MAP-KKK genes possess few intronless structures.

Gene expression is mainly used to determine their functional roles and plant physiological growth. Using the RNA sequencing data and RT-qPCR validation, expression patterns in seven *GHMAPK* genes were observed across vegetative and reproductive tissue in this study. The results show that all genes have been highly expressed across different tissues, suggesting these genes perform several functions during different phases of growth and development in cotton. Previous researches showed that specific MAPK genes have tissue-specific expression patterns in many plants such as cotton, wheat, maize, and cucumber (Chen and Murata 2011; Liu et al. 2013). However, these genes express across both vegetative as well as reproductive organs with different expression levels.

Both RNA sequence and RT-qPCR validation analyses indicated that *GHMPK* genes were upregulated across drought and salt stresses. Hence, results expressed that six genes were upregulated under drought treatment while only *GH_D11G0040* showed no expression, and six genes have been upregulated under salt treatment (*GH_A07G1527*, *GH_D02G1138*, *GH_D03G0121*, *GH_D03G1517*, *GH_D05G1003*, and *GH_D12G2528*) except *GH_D11G0040* with relatively low expression. This is consistent with previous findings that various abiotic stresses upregulated *GhMPK2* and *GbMPK3* in cotton and possibly enhanced drought and oxidative stress tolerance. The expression levels of most *GhMAPKs* have been found to increase in cotton when exposed to a water deficit, implying that these genes may be linked to cotton drought tolerance and response (Sadau et al. 2021; Teige et al. 2004; Zhang et al. 2011). Also, previous research confirmed the expression of the MAPK genes regulates drought stress across other plants such as rice, wheat, maize, and *Arabidopsis* (Lee et al. 2011; Rohila and Yang 2007; Zhan et al. 2017). MAPK genes have been reported to be induced through salt stress, as reported by previous researches (Teige et al. 2004; Xie et al. 2014). A study on genome-wide bioinformatics analysis of MAPK gene family in kiwifruit (*Actinidia chinensis*) confirmed that the expression levels of most *AcMAPK* gene were significantly higher at all treatment time points, whereas the

leftover genes are down-regulated in salinity stress. Also, studies confirmed that *TaMAPK29*, *TaMAPK33*, and *TaMAPK41* had been linked to induced salt stress (Wang et al. 2018).

Conclusion

The identification of *MAPK* genes from *Gossypium hirsutum* to enhance tolerance against drought and salt stresses based on family analysis of *G. hirsutum*, *G. arboreum*, and *G. raimondii*, has laid an underlying principle of genetic measures to cotton plant under drought and salt stress. *GHMAPK* elements related *cis*-regulatory elements analysis suggests that these genes significantly affect abiotic stress tolerance. Analysis of RNA-sequence and RT-qPCR data has revealed the upregulation of several genes across both vegetative and reproductive tissues. They have been declared candidate genes for tolerance to drought and salt stresses in cotton due to their upregulation across post-treatment examinations under the reported study. This research lays the groundwork to build a more robust cotton genotype that performs better under different environmental stress features, like drought and salt stress. Also, MAPK genes possess a wide role in regulatory properties and the extension of MAPK research is economically important to cotton crop production.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42397-022-00131-w>.

Additional file 1. Table S1: Physicochemical properties of MAPK genes

Additional file 2. Table S2: Functions of identified *cis*-regulatory element

Additional file 3. Table S3: Gene primers used for qRT-PCR

Author contributions

Sadau SB, Mehari TG, Ahmad A carried out the experiments and wrote the manuscript. Tajo SM, Elasad M, and Ibrahim S assisted in data collection. Iqbal MS, Zhang JL, Wei HL, and Yu SX revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by National Key R&D Program of China (2020YFD1001004).

Availability of data and materials

All the related data and files are all presented including the sequences of the primers used in the genes expression profiling.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declared that they have no competing interests.

Author details

¹State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, Henan 455000, China. ²Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan 430062, China. ³Department of Science and Technology Education, Bayero University Kano, P.O. Box 3011, Kano, Nigeria. ⁴Agricultural Research Corporation (ARC), P.O. Box 126, Wad Medani, Sudan.

Received: 15 November 2021 Accepted: 24 August 2022

Published online: 19 September 2022

References

- Bengough AG, McKenzie BM, Hallett PD, et al. Root elongation, water stress, and mechanical impedance: a review of limiting stresses and beneficial root tip traits. *J Exp Bot*. 2011;62(1):59–68. <https://doi.org/10.1093/jxb/erq350>.
- Bilas R, Szafran K, Hnatuszko-Konka K, Kononowicz A. *Cis*-regulatory elements used to 498 control gene expression in plants. *Plant Cell Tissue Organ Cult*. 2016;127:269–87. <https://doi.org/10.1007/s11240-016-1057-7>.
- Chen TH, Murata N. Glycinebetaine protects plants against abiotic stress: mechanisms and biotechnological applications. *Plant Cell Environ*. 2011;34:1–20. <https://doi.org/10.1111/j.1365-3040.2010.02232.x>.
- Chen C, Chen H, He Y, et al. TBtools, a Toolkit for biologists integrating various biological data handling tools with a user-friendly interface. *bioRxiv*. 2018. <https://doi.org/10.1101/289660>.
- Das K, Roychoudhury A. Reactive oxygen species (ROS) and response of anti-oxidants as ROS-scavengers during environmental stress in plants. *Front Environ Sci*. 2014;2:1–13. <https://doi.org/10.3389/fenvs.2014.00053>.
- Elasad M, Wei H, Wang H, et al. Genome-wide analysis and characterization of the TRX gene family in upland cotton. *Trop Plant Biol*. 2018;11:119–30. <https://doi.org/10.1007/s12042-018-9205-3>.
- Escobar-Sepúlveda HF, Go FC, Leo M. Expression patterns and promoter analyses of aluminum-responsive *NAC* genes suggest a possible growth regulation of rice mediated by aluminum, hormones and *NAC* transcription factors. *PLoS ONE*. 2017;12(10):1–25. <https://doi.org/10.1371/journal.pone.0186084>.
- Hamel LP, Nicole MC, Sritubtim S, et al. Ancient signals: comparative genomics of plant MAPK and MAPKK gene families. *Plants*. 2006;11(4):192–8. <https://doi.org/10.1016/j.plants.2006.02.007>.
- Hanin M, Brini F, Ebel C, et al. Versatile proteins for complex mechanisms plant dehydrins and stress tolerance. *Trends Plant Sci*. 2011;6(10):1503–9. <https://doi.org/10.4161/psb.6.10.17088>.
- Ichimura K, Mizoguchi T, Yoshida R, et al. Various abiotic stresses rapidly activate *Arabidopsis* MAP Kinases *ATMPK4* and *ATMPK6*. *Plant J*. 2000;24:655–65. <https://doi.org/10.1046/j.1365-313x.2000.00913>.
- Khan A, Pan X, Najeeb U, et al. Coping with drought: stress and adaptive mechanisms, and management through cultural and molecular alternatives in cotton as vital constituents for plant stress resilience and fitness. *Bio Res*. 2018;51(1):47. <https://doi.org/10.1186/s40659-018-0198-z>.
- Larade K, Storey KB. Analysis of signal transduction pathways during anoxia exposure in a marine snail : a role for p38 MAP kinase and downstream signaling cascades. *Comp Biochem Physiol*. 2006;143:85–91. <https://doi.org/10.1016/j.cbpb.2005.10.008>.
- Lee SK, Kim BG, Kwon TR, et al. Overexpression of the mitogen-activated protein kinase gene *OsMAPK33* enhances sensitivity to salt stress in rice (*Oryza sativa* L.). *J Biosci*. 2011;36(1):139–51. <https://doi.org/10.1007/s12038-011-9002-8>.
- Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. *Nucleic Acids Res*. 2002;30(1):325–7. <https://doi.org/10.1093/nar/30.1.325>.
- Li Y, Zhang L, Wang X, et al. Cotton *GhMPK6a* negatively regulates osmotic tolerance and bacterial infection in transgenic *Nicotiana benthamiana*, and plays a pivotal role in development. *FEBS J*. 2013;280:5128–44. <https://doi.org/10.1111/febs.12488>.
- Li YZ, Zhang L, Lu WJ, et al. Overexpression of cotton *GhMCK4* enhances disease susceptibility and affects abscisic acid, gibberellin and hydrogen peroxide signalling in transgenic *Nicotiana benthamiana*. *Mol Plant Path*. 2014;15(1):94–108. <https://doi.org/10.1111/mpp.12067>.
- Liu YK, Zhang D, Wang L, Li DQ. Genome-wide analysis of mitogen-activated protein kinase gene family in maize. *Plant Mol Biol Rep*. 2013;31:1446–60. <https://doi.org/10.1007/s11105-013-0623-y>.
- Long L, Gao W, Xu L, et al. *GbMPK3*, a mitogen-activated protein kinase from cotton, enhances drought and oxidative stress tolerance in tobacco. *Plant Cell Tissue Organ Cult*. 2014;116:153–62. <https://doi.org/10.1007/s11240-013-0392-1>.
- Lu P, Magwanga RO, Kirungu JN, et al. Genome-wide analysis of the cotton G-coupled receptor proteins (GPCR) and functional analysis of *GTOM1a*, a novel cotton GPCR gene under drought and cold stress. *BMC Genomics*. 2019;20:651. <https://doi.org/10.1186/s12864-019-5972-y>.
- Ma XJ, Yu TF, Li XH, et al. Overexpression of *GmNFYA5* confers drought tolerance to transgenic *Arabidopsis* and soybean plants. *BMC Plant Biol*. 2020;20:123. <https://doi.org/10.1186/s12870-020-02337-z>.
- Magwanga RO, Kirungu JN, Lu P, et al. Genome wide identification of the trihelix transcription factors and overexpression of *Gh_A05G2067 (GT-2)*, a novel gene contributing to increased drought and salt stresses tolerance in cotton. *Physiol Plant*. 2019;167(3):447–64. <https://doi.org/10.1111/pp1.12920>.
- Munns R, Tester M. Mechanisms of salinity tolerance. *Annu Rev Plant Biol*. 2008;59:651–81. <https://doi.org/10.1146/annurev.arplant.59.032607.092911>.
- Nakagami H, Pitzschke A, Hirt H, et al. Emerging MAP kinase pathways in plant stress signalling. *Trends Plant Sci*. 2005;10(7):339–46. <https://doi.org/10.1016/j.tplants.2005.05.009>.
- Page JT, Huynh MD, Liechty ZS, et al. Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3 Genes|Genomes|Genetics*. 2013;3(10):1809–18. <https://doi.org/10.1534/g3.113.007229>.
- Rohila JS, Yang Y. Rice mitogen-activated protein kinase gene family and its role in biotic and abiotic stress response. *J Integ Plant Biol*. 2007;49(6):751–9. <https://doi.org/10.1111/j.1672-9072.2007.00501>.
- Sadua SB, Ahmad A, Tajo SM, et al. Overexpression of *GhMPK3* from cotton enhances cold, drought, and salt stress in *Arabidopsis*. *Agronomy*. 2021;11:1049. <https://doi.org/10.3390/agronomy11061049>.
- Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C_T method. *Nat Protoc*. 2008;3(6):1101–8. <https://doi.org/10.1038/nprot.2008.73>.
- Shi J, Zhang L, An HL, et al. *GhMPK16*, a novel stress-responsive group D MAPK gene from cotton, is involved in disease resistance and drought sensitivity. *BMC Mol Biol*. 2011;12:22. <https://doi.org/10.1186/1471-2199-12-22>.
- Tamura K, Peterson D, Peterson N, et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28(10):2731–9. <https://doi.org/10.1093/molbev/msr121>.
- Teige M, Scheikl E, Eulgem T, et al. The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Mol Cell*. 2004;15:141–52. <https://doi.org/10.1016/j.molcel.2004.06.023>.
- Tian YF, Gu HH, Fan ZX, et al. Role of a cotton endoreduplication-related gene, *GaTOP6B*, in response to drought stress. *Planta*. 2019;249(4):1119–32. <https://doi.org/10.1007/s00425-018-3067-7>.
- Tuteja N. Abscisic acid and abiotic stress signaling. *Plant Sig Behav*. 2017;2(3):135–8. <https://doi.org/10.4161/psb.2.3.4156>.
- Wang NN, Zhao LL, Lu R, et al. Cotton mitogen-activated protein kinase4 (*GhMPK4*) confers the transgenic *Arabidopsis* hypersensitivity to salt and osmotic stresses. *Plant Cell Tissue Organ*. 2015;123:619–32. <https://doi.org/10.1007/s11240-015-0865-5>.
- Wang C, Lu WJ, He XW, et al. The cotton mitogen-activated protein Kinase Kinase 3 functions in drought tolerance by regulating stomatal responses and root growth. *Plant Cell Physiol*. 2016;57(5):1629–42. <https://doi.org/10.1093/pcp/pcw090>.
- Wang KB, Wendel JF, Hua JP. Designations for individual genomes and chromosomes in *Gossypium*. *J Cotton Res*. 2018;1:3. <https://doi.org/10.1186/s42397-018-0002-1>.
- Xie KB, Chen JP, Wang Q, et al. Direct phosphorylation and activation of a mitogen-activated protein kinase by a calcium-dependent protein kinase

- in rice. *Plant Cell*. 2014; 26(7):3077–89. <https://doi.org/10.1105/tpc.114.126441>.
- Zhan HS, Yue H, Zhao X, et al. Genome-wide identification and analysis of MAPK and MAPKK gene families in bread wheat (*Triticum aestivum* L.). *Genes*. 2017;8:284. <https://doi.org/10.3390/genes8100284>.
- Zhang L, Xi DM, Li SW, et al. A cotton group C MAP kinase gene, *GhMPK2*, positively regulates salt and drought tolerance in tobacco. *Plant Mol Biol*. 2011;77:17–31. <https://doi.org/10.1007/s11103-011-9788-7>.
- Zhang XT, Cheng TC, Wang GH, et al. Cloning and evolutionary analysis of tobacco MAPK gene family. *Mol Biol Rep*. 2013;40(2):1407–15. <https://doi.org/10.1007/s11033-012-2184-9>.
- Zhang J, Zou D, Li Y, et al. GhMPK17, a cotton mitogen-activated protein kinase, is involved in plant response to high salinity and osmotic stresses and ABA signaling. *PLoS ONE*. 2014a;9(4):1–12. <https://doi.org/10.1371/journal.pone.0095642>.
- Zhang X, Wang L, Xu X, et al. Genome-wide identification of mitogen-activated protein kinase gene family in *Gossypium raimondii* and the function of their corresponding orthologs in tetraploid cultivated cotton. *BMC Plant Biol*. 2014b;14(1):1–17. <https://doi.org/10.1186/s12870-014-0345-9>.
- Zhang JB, Wang XP, Wang YC, et al. Genome-wide identification and functional characterization of cotton (*Gossypium hirsutum*) MAPKKK gene family in response to drought stress. *BMC Plant Biol*. 2020;20:217. <https://doi.org/10.1186/s12870-020-02431-2>.

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

