Genome Wide Identification and Characterization of Mitogen Activated Protein Kinase (MAPK) Genes Reveals Their Potential in Enhancing Drought and Salt Stress Tolerance in Gossypium Hirsutum

Sadau Bello Salisu (✉ Sbsadau.ste@buk.edu.ng)  
Cotton Research Institute  
https://orcid.org/0000-0001-6298-7873

Teame Gereziher Mehari  
Cotton Research Institute

Adeel Ahmad  
Cotton Research Institute

Sani Muhammad Tajo  
Cotton Research Institute

Sani Ibrahim  
Oil Crops Research Institute Chinese Academy of Agricultural Sciences

Muhammad Shahid Iqbal  
Cotton Research Institute

Mohammed Elasad  
Agricultural Cooperative College: Agricultural Cooperative University

Jingjing Zhang  
Cotton Research Institute

Hengling Wei  
Cotton Research Institute

Shuxun Yu  
Cotton Research Institute

Research

Keywords: Cotton, MAPK gene, Drought stress, Salt stress and Genome-wide identification

Posted Date: December 15th, 2021

DOI: https://doi.org/10.21203/rs.rs-1078536/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

**Background:** Cotton crop is universally considered as protein and edible oil source besides the major contributor of natural fiber and is grown all around the globe. Unpredicted environmental stresses are becoming a significant threat to sustainable cotton production, ultimately leading to a substantial irreversible economic loss. Mitogen-activated protein kinase (MAPK), generally considered essential for recognizing environmental stresses through phosphorylating downstream signal pathways.

**Results:** In the current study, we have identified 74 MAPK genes across cotton, 41 from *G. hirsutum*, 19 from *G. raimondii*, whereas 14 have been identified through *G. arboreum*. The MAPK gene-proteins have been further interrogated to determine their physicochemical 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 clades. 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 stress.

**Conclusions:** RNA sequence and qPCR results represented 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 genes were upregulated across salt treatment.

**Background**

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

Several biotic and abiotic factors significantly impact cotton productivity (drought, heat, waterlogging, and salinity), causing significant losses in the focused agricultural sector productivity. Although breeding programs have made a positive attempt, old crop breeding techniques have limitations such as crossing barriers, long-time effects, and genetic disease transformation. The cotton plant self-stresses due to its
indeterminate growth habit; that is, it grows and expands before internal or external stresses impede growth and expansion (X. Zhang et al., 2014a).

The high amount of greenhouse gas emissions in the atmosphere and associated air pollution are significant causes of heatwaves, floods, and drought stress. Drought stress can significantly impact crop production, and the magnitude and length of the stress are also important factors. Availability of water is a critical factor in achieving long-term sustainability in crop production (Khan et al., 2018). Drought stress is a significant problem in cotton productivity because 50% global cotton supply comes from drought challenges. Cotton crops require improved yields and yield balance in both standard and moisture-stressed environments (Tuteja, 2007). Drought stress influences cotton plants’ growth and productivity by inducing several morpho-physiological 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).

High salinity is among the most significant environmental stress that plant experience. 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 & Tester, 2008; Zhu, 2020). 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. 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 (Dk et al., 2020).

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, transforms external stimuli into cells. Following three sequentially active kinases from MAPK cascades are MAP (MAP kinase 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 & Storey, 2006; Rohila & Yang, 2007). MAPKs are activated when plants are stressed (Li et al., 2014; Wang et al., 2016; Zhang et al., 2014). However, most MAPK research has been done on model plants like Arabidopsis and Tobacco (Li 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 like Arabidopsis thaliana AtMPK3 and AtMPK6 (Ichimura et al., 2000). MAPK functions in cotton have been studied in various ways, among them are the following: In GhMPK6, Abscisic acid induces catalase1 expression and H$_2$O$_2$ synthesis, osmotic stress, and bacterial infection are negatively regulated by GhMPK6a (Liu et al., 2013). A total of 74 MAPK genes with 41 across G.
**Materials And Methods**

**Plant materials, growth condition, and stress treatments**

TM-1, *G. hirsutum* L., as upland cotton, was used to evaluate tissue/organ expression. Plant for vegetative tissues was planted under 25 °C, with a light cycle of 16 hrs and 8 hrs dark cycle in a controlled chamber, while plants for reproductive tissues were harvested in the Cotton Research Institute Chinese Academy of Agricultural Sciences field, Anyang, Henan, China. Tissues like young leaves were collected at an early stage of planting; stems, true leaves, roots, and fibers were collected after two-week of planting from the cotton field. During 10 days post-anthesis, the flower was harvested from the field. To determine the GHMPK gene function in cotton under abiotic stresses, the CCRI10 variety was used for salt stress, while the H177 variety was used for drought stress. The cotton seedlings CCRI10 and H177 were planted and harvested under a laboratory-controlled growth chamber with a 25°C temperature and a 16/8 hrs light/dark cycle. A 15% polyethylene glycol (PEG-6000) induced drought stress treatment was applied to seedlings, and for salt stress, a 200 mM sodium chloride (NaCl) treatment was used (Y. Li et al., 2013; Ma et al., 2020). Sample collection was carried out at 0, 2, 4, 6, 12, 48, and 72 hours after treatment. Sampling for each treatment was carried out three times, and samples were immediately collected in liquid nitrogen and preserved under a temperature of -80 °C.

**Physiochemical properties analysis of MAPK genes in Cotton species**

Proteins encoded by MAPK in *Gossypium hirsutum* (tetraploid cotton genome), *Gossypium arboreum*, and *Gossypium raimondii* have been obtained online cotton database CottonFGD. (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 physiochemical characteristics of the cotton MAPK gene were retrieved from cotton FGD.

**Sequence alignments, phylogenetic tree construction and collinearity**
Three MPK. Cotton protein sequences were downloaded for *G. hirsutum*, and *G. arboreum*, and *G. raimondii*, through the online cotton database, cotton FGD. (http://www.cottonfgd.org/) and Phytozome (https://phytozome.jgi.doe.gov/) was followed for *A. thaliana*, sequences retrieval, and the neighbor-joining (NJ) approach to investigate the evolutionary relationship. Computer software package MEGA 7.0 (www.megasoftware.net) was used to construct a phylogenetic tree, considering Jones-Taylor-Thornton to be the substitution model through selecting 1000 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 significant were considered ≥ 90 significant. The Gene I.D.S., GFF3 files, and linked files were used to construct the collinearity using the TBtools software.

**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 MAPK gene-related conserved motifs. TB tools software was then used for the motif visualization. The coding sequences (CDSs) were compared with the MAPK gene's genome sequences through an online gene structure tool (http://gsds.cbi.pku.edu.cn/). 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 2000bp distance from the start codon has been considered and searched for CARE program. (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/ (Lescot et al., 2002).

**Subcellular localization**

For localization of GHPMK protein, the protein sequence was downloaded from cottonFGD, and the prediction was carried out through Wolf PSORT (https://wolfsort.hgc.jp) online server.

**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 & Polyphenolics-rich) (Tiangen, China). Using a PrimeScriptTMRT reagent package with a gDNA Eraser, RNA has been further transcribed to cDNA through using rtqPCR as outlined. For calculation of relative expression, $2^{-\Delta\Delta CT}$ method has been followed (Schmittgen & 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**

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

Results

Physiochemical traits of MAPK gene family

A total of 74 MAPK genes were detected in cotton; out of them, 41 in Gossypium hirsutum, 19 in Gossypium raimondii, whereas 14 were Gossypium arboreum. These MAPK genes' proteins were studied further to evaluate their physicochemical characteristics as well as other features. These MAPK genes were reported to express proteins ranging in length between 604 to 87 amino acids, with molecular mass between 71.503 to 9.988 KDa and Isoelectric Points (pl) between 10.23 to 5.672. All identified MAPK genes have grand average hydropathy > 0, implying that all MAPK genes in cotton were hydrophilic (Table S2.1). Proteins encoded by stress-responsive genes are closely correlated with hydrophilicity (Magwanga et al., 2019). Hydrophilic associated genes were discovered to play several essential roles across various plants, including responses to stresses (Lu et al., 2019).

Cotton MAPK gene phylogenetic study with other plants

A phylogenetic tree was constructed based on multiple sequence alignment of 41 GHMAPK, 14 GaMPK protein sequences, 19 GrMPK protein sequences, and 18 Arabidopsis MAPK protein sequences revealed 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 2.1). TDY phosphorylation site may comprise all the genes in clades A, TEY motif across clade B members, Clade C members may contain the TDY motif, D members contain more TDY motif, and few TEY motifs lastly, 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 contains the TEY motifs. This shows that the cotton MAPK gene has more TDY motif and minor TEY motif while the AtMPK has more TEY motif and minor TDY motif. The cotton MAKS classification was consistent with previous findings (X. Zhang et al., 2014b). This indicates that the TEY MAPKs motif may significantly function in dicot plants than the TDY MAPKs motif.

To differentiate the collinear gene pairs, collinearity was done for the MAPK in three cotton species through circle gene viewer using the software package TBtools (C. Chen et al., 2018b). Collinearity analysis was performed between the physical maps of the subgenomes about Gossypium hirsutum, Gossypium raimondii, and Gossypium arboreum for associations among A vs. D, A vs. At, and D Vs. Dt subgenome. 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 genes structure analysis of MAPK

Investigations for conserved motif about MAPK KK 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 GHMAKS 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, 16, 16, and 19) and only 3 genes; GH_A11G0035, GH_A12G2888, GH_D11G0040, and GH_D12G2911 contains motif 20. While in GaMAPK majority of identified gene contains 9 similar motif composition (1, 2, 3, 4, 6, 8, 9, 10, 15) few among them contains 5 motifs (5, 7, 16, 17, 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 8 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, 20) and only few among the gene; Gorai.008G065400, Gorai.011G1006001 and Gorai.003G012800.1 contain motif 14.

Gene structure was examined by using an online tool http://gsds.cbi.pku.edu.cn/ for cotton MAPK KKs. In G. hirsutum, out of 41 genes, 38 were found to possess intron, and 3 are intronless. The most extended intron interruption was observed in GH_A11G0035 and GH_A12G2888 (Fig. 4). In G. arboreum, all the 19 genes possess intron, and the highest intron interruption was observed in Ga02G0944 (Fig. 4B), also in G. raimondii, both genes possess intron, and the highest intron interruption is found in Gorai.007G004400 and Gorai.008G249800 with 10 introns respectively (Fig 4C).

Chromosomal mapping of MAPK Genes in Gossypium species

The chromosome distribution was investigated in the 3 species of cotton and found out that among the 41 G. hirsutum MAPK genes, 20 were located in the At-sub genome, and 21 were located at the Dt subgenome. With four genes each, chromosome A12 and its homologous D12 had the highest gene locus, followed by chromosome A05. Three genes were found in A01, A03, D03, and D05, D07, 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 in A01, A02, A03,05, A11, and A12. The highest gene locus was observed in chromosomes A03, A11, and A12, with three genes (Fig. 5B). While A01, A02, and A05 both possess two genes, respectively. 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, D02,05, 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 (Bilas 2016). We have analyzed Cotton MAPKs promoter regions for the determination of their cis-elements. The 2kb sequences from the start of transcription from each
The cis-elements identified in the GHMPK promoter region were categorized into five functions: hormone responsiveness, stress responsiveness, light responsiveness, cellular responsiveness, and binding site (Table S 2.2). The majority of GHMPK 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 related to abiotic and biotic stresses in plants have been found in the promoter sequences in the coding sequence related to the GHMAPK gene, indicating that this gene can be investigated as an abiotic tolerance gene in cotton.

Subcellular analysis for MPK proteins in *Gossypium hirsutum*

Based on the WOLF PSORT ([https://wolfpsort.hgc.jp](https://wolfpsort.hgc.jp)) analysis result, GHMPK proteins are localized across different parts of the cell, including the nucleus cytoplasm, mitochondria, chloroplast Peristomes cytoskeleton, and Golgi-apparatus. The proteins were predominantly located on the cytoplasm (70%), nucleus (20%), and cytoskeleton (5%).

RNA Sequence Analysis and RT-qPCR confirmation of GHMPK 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 qRT-PCR primers were designed (Table S23). Based on RNA-sequence data as well as qRT-PCR results, we determined that GH_A07G1527 and GH_D02G1138 are upregulated in all tissues GH_D05G1003 are upregulated in all the tissues. Lastly, GH_D11G0040 shows no expression (Fig. 7A). The results demonstrated that RNA Sequence analysis and RT-qPCR expression results represented a strong correlation, with R² = 0.91 in drought, R² = 0.75 in salt and R² = 0.66 in tissues.

To determine the roles of GHMAPK genes across salt and drought stresses, the seven genes have been analyzed for both the RNA-sequence data and qRT-PCR results to detect their expression pattern after treatment. For drought treatment, GH_A07G1527 was upregulated at 6 hr and 12 hr, GH_D02G1138 was upregulated at 2 hr, 6 hr, and 12 hr, GH_D03G0121 was upregulated at 12 hr. At the same time, GH_D03G1517 and GH_D02G1138 were upregulated at 2 hr, 6 hr, and 12 hr. While GH_D11G0040 showed no expression. Lastly, GH_D12G2528 was upregulated at 12 (Fig. 7B). For salt treatment, GH_A07G1527 and
were upregulated at 2 hr, 6 hr, and 12 hr, GH_D03G0121 show no expression in RNA seq data and upregulated expression at 12hr post-treatment. GH_D03G1517 and GH_D02G1138 were upregulated at 2hr, 6hr, and 12hr of post-treatment, respectively. GH_D11G0040 showed no expression in RNA seq data and upregulated at 6 hr and 12 hr of qRT-PCR. Lastly, GH_D12G2528 at 2hr, 6hr, and 12h respectively (Fig. 7C).

**Discussion**

Plants are constantly vulnerable to different environmental stress conditions, for example, pathogen infections, salt, cold, drought, and even oxidative stress. Such environmental stresses have adverse effects on plant development and productiveness, resulting from significant loss of crop productivity (Tuteja, 2007). 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 (C. 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 (N. W. Li et al., 2015; Long et al., 2014; Sadau et al., 2021; Shi et al., 2011; Jie Zhang et al., 2014; Jing-bo Zhang et al., 2020; L. Zhang et al., 2011). To date, more studies are needed to strengthen knowledge about the sophisticated biological functions with particular reference to MAPK cascades in cotton.

About 41 MAPK genes in *G. hirsutum*, 15 in *G. arboreum*, whereas 19 in *G. raimondii* have been found. 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. Previous research has confirmed that hydrophilic proteins are involved in tolerating numerous abiotic stresses (Hanin et al., 2011). Cotton has more MAPKs genes in Clade A than Clade B, according to the results of the phylogenic 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). Prevalence of such elements across the promoter region of these genes signifies their role in the growth and development of plants (Das & Roychoudhury, 2014; Elasad et al., 2018; Escobar-sepu et al., 2017).

Analysis of exon/ intron structure revealed that most of the Cotton MAPK gene was interrupted by an intron. Few were intronless. The longest intron in *G. hirsutum* is found in GH-A11G0035 and
Gene expression is mainly used to determine their functional roles and plant physiological growth. Using the RNA sequencing data and RT-qPCR validation, observation of expression patterns in seven GHMAPK genes across vegetative and reproductive tissue in this study. The result shows 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 research showed that specific MAPK genes have a tissue-specific expression in many plants such as cotton, wheat maize, and cucumber (T. H. H. Chen & Murata, 2011; Liu et al., 2013). However, these genes have expression across both vegetative as well as reproductive organs with different expression levels.

Both RNA sequence and RT-qPCR validation Analyses indicated that GHMPK genes upregulated across drought and salt stresses. Hence, results expressed that all the six genes were upregulated under drought treatment, whereas all the seven genes have been upregulated under salt treatment, and two were downregulated. This is consistent with previous findings that various abiotic stresses upregulate GhMPK2 and GbMPK3 in cotton and possibly enhanced drought and oxidative stress tolerance. The expression of most GhMAPKKKs has 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; L. Zhang et al., 2011). Also, previous researches confirmed the expression of the MAPK genes across other plant species, including Rice, Wheat, Maize, and Arabidopsis (Lee et al., 2011; Rohila & Yang, 2007; Triticum, n.d.). For salt treatment, 3 among the genes were upregulated at various hours after treatment, one is upregulated at 48 hr, and the rest of the 3 genes were downregulated. MAPK genes have been reported to be induced through salt stress, as reported by previous researches (Teige et al., 2004; Xie et al., 2014). Research studies by Wang et al., 2018 that most AcMAPK gene expressions, were significantly higher at all treatment time points, whereas the leftover genes are down-regulated in salinity stress. Also, research studies confirmed that TaMAPK29, TaMAPK33, and TaMAPK41 had been linked to induced salt stress.

Conclusion

Environmental challenges remain critical in crop production, even though significant measures were made to control genetic mechanisms underlying abiotic stress tolerance. Thus water availability, temperature maintenance, and disease control are of paramount importance. Crops are generally exposed to multiple stresses, and hence the area that requires much more attention is the plant response to these stresses. These require an integrated approach to water-deficient lands, mining critical genes to enhance stress tolerance through conventional breeding methods. This was executed to investigate MAPK genes from Gossypium hirsutum to enhance tolerance against drought, and salt stresses based on family analysis, Gossypium hirsutum, Gossypium arboreum, and Gossypium raimondii, respectively. GHMAPK elements related cis-regulatory elements analysis suggests that these genes significantly affect
abiotic stress tolerance. Analysis of RNA-sequence and RT-qPCR data have revealed the upregulation of several genes across both vegetative and reproductive tissue. 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 work lays the groundwork for more research into these genes to build a more robust cotton genotype that performs better under different environmental stress features, like drought, cold and salt stress.

Declarations

Author Contributions

Sadau SB, Mehari TG, Ahmad A conducted the experiment and wrote the manuscript. Tajo SM, Elasad M, and Ibrahim S, assisted in data collection. Iqbal MS, Zhang J, Wei H, and Yu S revised the manuscript. All authors reread and agreed on the final manuscript.

Funding

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

Availability of data and materials

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

Ethics approval and consent to participate. No ethical nor consent to contribute in this research was sought, this not application in this research work.

Consent for publication

Not applicable.

Competing interests

The authors declared that they have no competing interests.

Author details

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

1. Bengough, A. G., McKenzie, B. M., Hallett, P. D., & Valentine, T. A. Root elongation, water stress, and mechanical impedance: a review of limiting stresses and beneficial root tip traits. *Journal of Experimental Botany*, 2011; 62(1), 59–68. https://doi.org/10.1093/jxb/erq350

2. Bilas 2016.pdf. (n.d.).

3. Chen, C., Chen, H., He, Y., & Xia, R. TBtools, a Toolkit for Biologists integrating various biological data handling tools with a user-friendly interface. *BioRxiv*. 2018a; https://doi.org/10.1101/289660

4. Chen, C., Chen, H., He, Y., & Xia, R. TBtools, integrating various biological data handling; with a user. *BioRxiv*, 289660. 2018b. https://doi.org/10.1101/289660

5. Chen, T. H. H., & Murata, N. *Glycinebetaine protects plants against abiotic stress: mechanisms and biotechnological applications*. 2011;1–20. https://doi.org/10.1111/j.1365-3040.2010.02232.x

6. Das, K., & Roychoudhury, A. *Reactive oxygen species (ROS) and response of antioxidants as ROS-scavengers during environmental stress in plants*. 2014;2, 1–13. https://doi.org/10.3389/fenvs.2014.00053

7. Dk, S., Manikandan, A., Blaise, D., & Shukla, P. K. *Identification of Relationship among Exogenous NaCl with Cotton Leaves on Cation Uptake, Nutrient Ratios and Status in Rhizosphere Soil*. 2020;9 (36), 956–965. https://doi.org/10.37273/chesci.CS205110221

8. Elasad, M., Wei, H., Wang, H., Su, J., Ondati, E., & Yu, S. Genome- Wide Analysis and Characterization of the TRX Gene Family in Upland Cotton. *Tropical Plant Biology*, 2018;11(3–4), 119–130. https://doi.org/10.1007/s12042-018-9205-3

9. Escobar-sepu, H. F., Go, F. C., & 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*. 2017;1–25. https://doi.org/10.1371/journal.pone.0186084

10. Hamel, L., Nicole, M., Sritubtim, S., Ellis, M., Ehlting, J., Beaudoin, N., Barbazuk, B., Klessig, D., Lee, J., Martin, G., Mundy, J., Ohashi, Y., Scheel, D., Sheen, J., Xing, T., Zhang, S., Seguin, A., & Ellis, B. E. *Ancient signals: comparative genomics of plant MAPK and MAPKK gene families*. 2006;11(4). https://doi.org/10.1016/j.tplants.2006.02.007

11. Hanin, M., Brini, F., Ebel, C., Toda, Y., Takeda, S., & Masmoudi, K. *Versatile proteins for complex mechanisms Plant dehydrins and stress tolerance*. 2011;1503–1509. https://doi.org/10.4161/psb.6.10.17088

12. Ichimura, K., Mizoguchi, T., Yoshida, R., Yuasa, T., & Shinozaki, K. (). *Various abiotic stresses rapidly activate Arabidopsis MAP*. 2000;24, 655–665.

13. Khan, A., Pan, X., Najeeb, U., Tan, D. K. Y., Fahad, S., Zahoor, R., & Luo, H. 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. *Biological Research*, 2018;51(1), 47. https://doi.org/10.1186/s40659-018-0198-z
14. Kunbo, W., F, W. J., & Jinping, H. U. A. *Designations for individual genomes and chromosomes in Gossypium* 2018; 3–7.

15. 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.* 2006;143, 85–91. https://doi.org/10.1016/j.cbpb.2005.10.008

16. 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.*). *Journal of Biosciences,* 2011;36(1), 139–151. https://doi.org/10.1007/s12038-011-9002-8

17. 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 Research,* 2002;30(1), 325–327. https://doi.org/10.1093/nar/30.1.325.

18. Li NW, Zhao L, Lu R, & Li Y. Cotton mitogen-activated protein kinase4 (GhMPK4) confers the transgenic Arabidopsis hypersensitivity to salt and osmotic stresses. *Plant Cell, Tissue and Organ Culture (PCTOC),* 2015;4. https://doi.org/10.1007/s11240-015-0865-5

19. Li Y, Zhang L, Lu W, et al. *Overexpression of cotton GhMKK4 enhances disease susceptibility and affects abscisic acid, gibberellin and hydrogen peroxide signalling in transgenic Nicotiana benthamiana.* 2014; 15, 94–108. https://doi.org/10.1111/mpp.12067

20. 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.* 2013; 280, 5128–5144. https://doi.org/10.1111/febs.12488

21. Liu Y, Zhang D, Wang L, & Li D. *Genome-Wide Analysis of Mitogen-Activated Protein Kinase Gene Family in Maize.* 2013; https://doi.org/10.1007/s11105-013-0623-y

22. Long L, Gao W, Xu L, & Liu M. *GbMPK3, a mitogen-activated protein kinase from cotton, enhances drought and oxidative stress tolerance in tobacco.* 2014; 153–162. https://doi.org/10.1007/s11240-013-0392-1

23. Lu P, Magwanga RO, Kirungu JN. et al. *Genome-wide analysis of the cotton G- coupled receptor proteins (GPCR) and functional analysis of GTOM1, a novel cotton GPCR gene under drought and cold stress.* 2019; 1–17.

24. Ma X, Yu T, Li X, et al. *Overexpression of GmNFYA5 confers drought tolerance to transgenic Arabidopsis and soybean plants.* 2020; 1–18.

25. Magwanga RO, Kirungu JN, Lu P, Yang X, & Dong Q. *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.* 2019; 2067, 447–464. https://doi.org/10.1111/ppl.12920

26. Munns R, & Tester M. *Mechanisms of Salinity Tolerance.* 2008; https://doi.org/10.1146/annurev.arplant.59.032607.092911

27. Nakagami H, Pitzschke A, & Hirt H et al. *Emerging MAP kinase pathways in plant stress signalling.* 2005; 10(7). https://doi.org/10.1016/j.tplants.2005.05.009
28. Page JT, Huynh MD, Liechty ZS et al. Insights into the evolution of cotton diploids and polyploids from whole-genome re-sequencing. *G3 (Bethesda, Md.)*, 2013; 3(10), 1809–1818. https://doi.org/10.1534/g3.113.007229

29. Rohila JS, & Yang Y. *Rice Mitogen-activated Protein Kinase Gene Family and Its Role in Biotic and Abiotic Stress Response.* 2007; 49(6), 751–759. https://doi.org/10.1111/j.1672-9072.2007.00501.x

30. Sadau SB, Ahmad A, Tajo, SM Ibrahim et al., *Overexpression of GhMPK3 from Cotton Enhances Cold, Drought, and Salt Stress in Arabidopsis.* 2021; 1–18.

31. Schmittgen TD, & Livak KJ. *Analyzing real-time PCR data by the comparative C T method.* 2008; 3(6), 1101–1108. https://doi.org/10.1038/nprot.

32. Shi J, Zhang L, An H, et al. *GhMPK16, a novel stress-responsive group D MAPK gene from cotton, is involved in disease resistance and drought sensitivity.* 2011.

33. Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*, 2011; 28(10), 2731–2739. https://doi.org/10.1093/molbev/msr121

34. Teige M, Scheikl E, Eulgem T, et al. *The M KK2 Pathway Mediates Cold and Salt Stress Signaling in Arabidopsis.* 2004; 15, 141–152.

35. Tian Y, Gu H, Fan Z, et al. Role of a cotton endoreduplication-related gene, GaTOP6B, in response to drought stress. *Planta*, 2019; 249(4), 1119–1132. https://doi.org/10.1007/s00425-018-3067-7

36. Triticum, L. (n.d.). *and MAPKK Gene Families in Bread Wheat.* https://doi.org/10.3390/genes8100284

37. Tuteja N, *Abscisic Acid and Abiotic Stress Signaling.* 2007; June, 135–138.

38. Wang C, Lu W, He X, et al. *The Cotton Mitogen-Activated Protein Kinase Kinase 3 Functions in Drought Tolerance by Regulating Stomatal Responses and Root Growth.* 2016; 57(May), 1629–1642. https://doi.org/10.1093/pcp/pcw090

39. Wang F, Wang C, Yan Y. et al. *Overexpression of Cotton GhMPK11 Decreases Disease Resistance through the Gibberellin Signaling Pathway in Transgenic Nicotiana benthamiana.* 2016; 7(May), 1–16. https://doi.org/10.3389/fpls.2016.00689

40. Wang G, Wang T, Jia H, et al. Genome-wide bioinformatics analysis of MAPK gene family in kiwifruit (Actinidia chinensis). *International Journal of Molecular Sciences*, 2018; 19(9). https://doi.org/10.3390/ijms19092510

41. Wang G, Wang T, Jia Z, *Genome-Wide Bioinformatics Analysis of MAPK Gene Family in Kiwifruit (Actinidia Chinensis).* 2018; https://doi.org/10.3390/ijms19092510

42. Xie K, Chen J, Wang Q, et al. *Direct Phosphorylation and Activation of a Mitogen-Activated Protein Kinase by a Calcium-Dependent Protein Kinase in Rice.* 2014; 26(July), 3077–3089. https://doi.org/10.1105/tpc.114.126441

43. Zhang J, Zou D, Li Y. *GhMPK17, a Cotton Mitogen-Activated Protein Kinase, Is Involved in Plant Response to High Salinity and Osmotic Stresses and ABA Signaling.* 2014; 9(4), 1–12.
44. Zhang J, Wang X, Wang Y, Genome-wide identification and functional characterization of cotton (Gossypium hirsutum) MAPKKK gene family in response to drought stress. 2020; 1–14.

45. Zhang L, Xi D, & Li S, et al A cotton group C MAP kinase gene, GhMPK2, positively regulates salt and drought tolerance in tobacco. 2011; 17–31. https://doi.org/10.1007/s11103-011-9788-7

46. 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 Biology*, 2014a; 14(1), 345. https://doi.org/10.1186/preaccept-3285223921360401

47. 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 Biology*, 2014b; 14(1), 1–17. https://doi.org/10.1186/s12870-014-0345-9

48. Zhu J. *Update on Stress Signaling Genetic Analysis of Plant Salt Tolerance Using Arabidopsis*. 124 2020; 941–948.

**Figures**

**Figure 1**

MEGA 7 was used for the development of the phylogenetic tree. Different clades of MAPKs are represented by Letters A-E. MAPKs are genes highlighted in different colors, G. hirsutum highlighted in red, G. arboreum highlighted in cadet blue while G. raimondii highlighted in violet and A. thaliana highlighted in green

**Figure 2**

Collinearity Analysis of MAPK genes in cotton. Dark green: G. arboreum; Red: G. raimondii, Light green: At and Dt of G. hirsutum

**Figure 3**

Gene motif (A) Gossypium hirsutum (B) Gossypium raimondii (C) Gossypium arboreum

**Figure 4**
Analysis of Gene structure using Gene structure display server (A) Gossypium hirsutum (B) Gossypium raimondii

Figure 5

Chromosomal mapping of Cotton MAPK Genes. GHMAPK genes were Mapped on (A). At subgenome (B). Dt subgenome (C). GaMAPK were mapped A1, A2, A3, A5, A11, and, A12 (D) GrMAPK were mapped on D1, D2, D3, D5, D6, D7, D8, D9 and D11 chromosome. respectively. Chromosome number is indicated on the middle of the left-hand side of the chromosome.

Figure 6

Analysis of Gene structure using Gene structure display server (A) Gossypium hirsutum (B) Gossypium raimondii (C) Gossypium arboreum.

Figure 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.

Supplementary Files

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

- TableS1.docx
- TableS2.docx
- TableS3.docx