

Association analysis of GmMAPKs and functional characterization of GmMMK1 to salt stress response in soybean

Xiliang Liao¹  | Meiqi Shi¹ | Wei Zhang^{1,2} | Qian Ye¹ | Yali Li¹ |
Xianzhong Feng³ | Javaid Akhter Bhat¹  | Guizhen Kan¹  | Deyue Yu^{1,4}

¹National Key Laboratory of Crop Genetics and Germplasm Enhancement, National Center for Soybean Improvement, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing, China

²Institute of Industrial Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China

³Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China

⁴School of Life Sciences, Guangzhou University, Guangzhou, China

Correspondence

Guizhen Kan and Deyue Yu, National Key Laboratory of Crop Genetics and Germplasm Enhancement, National Center for Soybean Improvement, Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing, China.
Email: kanguizhen@njau.edu.cn and dyu@njau.edu.cn

Funding information

EUCLEG Horizon 2020 of European Union, Grant/Award Number: 727312; Ministry of Science and Technology, Grant/Award Number: 2017YFE0111000; Natural Science Foundation of Jiangsu Province, Grant/Award Number: BK20191313

Edited by R. Rivero

Abstract

Salinity is one of the major abiotic constraints affecting the growth and yield of plants including soybean. In this context, the previous studies have documented the role of the mitogen-activated protein kinase (MAPK) cascade in the regulation of salt signaling in model plants. However, there is not a systematic analysis of salt-related MAPKs in soybean. Hence, in this study, we identified a total of 32 GmMAPKs via genome-wide reanalysis of the MAPK family using the soybean genome v4.0. Based on the transcriptome datasets in the public database, we observed that GmMAPKs are induced by different abiotic stresses, especially salt stress. Furthermore, based on the candidate gene association mapping and haplotype analysis of the GmMAPKs, we identified a salt-related MAPK member, GmMMK1. GmMMK1 possesses significant sequence variations, which affect salt tolerance in soybean at the germination stage. Besides, the over-expression of the GmMMK1 in soybean hairy roots has a significant negative effect on the root growth, leading to increased sensitivity of the GmMMK1-OE plants to salt stress. Moreover, the heterologous expression of the GmMMK1 in *Arabidopsis* has been also observed to have a negative effect on the germination and root growth under salt stress. The transcriptome analysis and yeast two-hybrid screening showed that hormone signaling and the homeostasis of reactive oxygen species are involved in the GmMMK1 regulation network. In conclusion, the results of this work demonstrated that GmMMK1 is an important negative regulator of the salt stress response, and provides better insights for understanding the role of the MAPKs in soybean salt signaling.

1 | INTRODUCTION

Plants being sessile are often subjected to environmental stresses that negatively affect their growth and development (Bhat et al., 2020). In this regard, soil salinity has become a severe problem of agricultural crop productivity worldwide. Salt stress directly affects the germination, plant

growth, and final yields (Isayenkov & Maathuis, 2019). The ability of the plant to germinate under salt stress conditions is an important component of salt tolerance (ST). Soybean [*Glycine max* (L.) Merr.] is a distinctive food and cash crop; however, this crop is highly sensitive to salt stress (Farooq et al., 2017). First, the germination of soybean seeds could be delayed under low-salt conditions (below 0.1% NaCl). What's worse, the germination rate gradually and significantly decreases under higher concentrations of salt stress (Phang et al., 2008). In previous studies, we

Xiliang Liao and Meiqi Shi are joint first authors.



demonstrated that the germination rate of the diverse soybean germplasms under salt stress fits a normal continuous distribution, indicating that ST at the germination stage is a complex quantitative trait controlled by multiple genes (Kan et al., 2015, 2016). Furthermore, relative germination rate can be used as the phenotypic trait to identify new major ST related genes, such as *GmCDF1* (Zhang et al., 2019).

MAPK cascade is a highly conserved signaling module acting downstream of the receptors that transduce extracellular stimuli into intracellular responses in eukaryotes (Meng & Zhang, 2013). Recently, there is a big progression of knowledge concerning the MAPK cascade as an early key component of osmotic signaling (Fàbregas et al., 2020). During the soybean germination stage, it has been revealed that osmotic, as well as oxidative stress, is the major adaptive response to salinity (Fercha et al., 2016). However, few studies have suggested the relationship between MAPK cascade genes and ST in soybean germination or seedlings stage. By using the comparative proteomic analysis, a soybean MAPK member named MMK2 was identified as a significantly differentially expressed protein under salt treatment (Ji et al., 2016). Similarly, using RNA sequencing (RNA-seq) analysis, the members of MAPK cascade were observed to be enriched under salt stress treatment (Liu et al., 2015; Wang et al., 2018). It seems that no matter the protein level or transcriptional level, MAPK cascade members responded quickly and critically to salt stress. Furthermore, by using the GWAS analysis, *Glyma.18g47140* (an *Arabidopsis* MPK4 homolog) of soybean was identified as a candidate gene regulating ST in soybean (Kan et al., 2015). However, these studies did not carry out detailed research on the MAPK gene family or a specific salt-related member.

With the availability of the reference genome sequence of different plant species, the genome-wide characterization and identification of gene families have become routine work. The *Arabidopsis* genome encodes approximately 60, 10, and 20 members of MAP3Ks, MAP2Ks, and MAPKs, respectively (MAPK Group, 2002). The *Arabidopsis* MAPKs had been divided into four groups, and each group functions specifically in the plant growth or stress regulation; for example, the group A members are more involved in development processes, while group B members such as AtMPK4s participate in pathogen defense and abiotic stress (Andreasson & Ellis, 2010). After the first soybean genome was released (Schmutz et al., 2010), a total of 38 MAPKs, 11 MAPKKs, and 150 MAPKKKs in the soybean genome were identified soon (Neupane et al., 2013). However, with the promotion of the soybean genome assembly and the improvement of gene function annotation in model plants, the classification and precise annotations of the soybean MAPKs show more importance but its progress is quite slow. Moreover, it is very meaningful to find one representative member or subfamily in MAPKs that is responsible for the regulation of ST. Last but not least, the constant challenge in understanding MAPK signaling pathways is the difficulty in identifying connections between kinase activation and downstream effector proteins (Pitzschke, 2015).

By keeping the above into consideration, the present study using multi-omics analysis like genomic variations and transcriptomes to finally identify a model of salt-related MAPKs, *GmMMK1*. Then, we elucidated the negative role of the *GmMMK1* in the salt signaling of

soybean and what we found may provide new insights into the role of MAPKs in early salt signaling.

2 | MATERIALS AND METHODS

2.1 | Identification of putative MAPKs in soybean

The 38 members of the MAPKs have been characterized at the genome-wide level by using the previous soybean draft of the whole genome sequence as reference (Neupane et al., 2013), while an updated version of the soybean whole sequence draft was released recently (Valliyodan et al., 2019). Thus, we reanalyzed the MAPK gene family by using the updated version of the soybean reference genome sequence. The latest released soybean genome assembly version v4.0 and annotation version Wm82.a4.v1 were downloaded from the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>). We retrieved the sequence alignment of two conserved MAPK domains, namely, stkc tdy mapk and stkc tey mapk, from the NCBI Conserved Domain Database (CDD; <https://www.ncbi.nlm.nih.gov/cdd/>). The two domains were searched in the soybean genome using HMMER 3.3 with an e-value cutoff of 1×10^{-50} . At the same time, we retrieved 20 *Arabidopsis* MAPKs from the TAIR (<https://www.arabidopsis.org/>) and 17 rice MAPKs from UniProt (<https://www.uniprot.org/>) as guidance criteria for a BLASTP algorithm; and these known MAPKs of the *Arabidopsis* and rice were used as query. The cutoff value for this was also set as 1×10^{-50} . Finally, a keyword search using “MAPK” was performed in the following protein databases: NCBI Gene, UniProt, and Pfam to do a manual check.

2.2 | Phylogenetic and structural analysis of MAPKs

The alignment of the MAPK sequences was performed through the CLUSTALW alignment function in MEGA 7.0 (Kumar et al., 2016). Phylogenetic tree was constructed by using the maximum likelihood method and 1000 bootstraps values. MAPKs of soybean along with known MAPKs of *Oryza sativa*, and *Arabidopsis* were used for the construction of the phylogenetic tree.

MEME program was employed for the detection of conserved motifs in all protein sequences (Bailey et al., 2009). Following parameters were set in the MEME analysis: the maximum number of motifs was set to 15; motif width was set to ≥ 6 and ≤ 200 ; motif repetition was set to zero or one occurrence per sequence. The identified motifs were annotated using the InterProScan program. Conserved domains in the MAPKs were identified by using the NCBI conserved domains database (CDD).

2.3 | Expression pattern of MAPKs under multiple environmental stresses

Freely available expression data of MAPKs were obtained online from the SoyBase database (<https://www.soybase.org/soyseq/>). The

collection of soybean transcriptomes related to environmental stress was based on the plant expression database PLEXdb (<https://www.plexdb.org/>). By using the GEO2R function from NCBI (<https://www.ncbi.nlm.nih.gov/geo/>), we obtained the corresponding expression profiles of microarray datasets. The RNA-seq data of a time-course experiment under salt treatment were reanalyzed (Liu et al., 2019). Gene expression omnibus (GEO) accessions of all expression data used are shown in Table S2.

2.4 | Quantitative RT-PCR for determining the expression level of *GmMMK1*

To analyze the expression of *GmMMK1* under salt stress, we chosen two distinct materials in ST viz., NJAU_C002 and NJAU_C101. Their seeds were initially sterilized, then put into the sterilized water to make the seeds imbibed, and treated with 0 or 150 mM NaCl. After treatment with or without salt stress, the samples that is, soybean embryos were collected at the 2, 4, 24, and 48 h time intervals. As for the seedling samples, the 2-week-old seedlings' roots were immersed in the 1/2 Hoagland's solution saturated with 150 mM NaCl, and then sampled at the same time points. Gene expression was determined by RT-PCR analysis using an ABI 7500 system (Applied Biosystems) with the SYBR Green Real-time Master Mix (Vazyme), and the data were analyzed using ABI 7500 Sequence Detection System (SDS) software version 1.4.0. The normalized expression was calculated for each sample as $\Delta\Delta CT$, and the fold change was calculated as $2^{-\Delta\Delta CT}$, which represents the relative expression level of *GmMMK1* under different salt treatment time points.

2.5 | Plant materials and evaluation of salt tolerance (ST)

In the present study, we used the natural population consisting of 219 cultivated soybean accessions for the trait-marker association analysis. Seeds used in the germination experiments were collected from two different environments viz., Nanjing, China, in 2013 (E1) and Nantong, China, in 2017 (E2).

Healthy and viable seeds of each soybean accession were selected for surface sterilization for 6 h using the chlorine gas, and were washed thoroughly with distilled water. The detailed experimental procedure was followed as described previously (Kan et al., 2015). Forty uniform seeds were then placed on two sheets of filter paper in sterilized Petri dishes, and treated with sterilized water or 150 mM NaCl solution. The seeds were incubated in a growth chamber at $25 \pm 1^\circ\text{C}$ in the dark for 7 days. The number of germinated seeds was counted to calculate the germination rate (GR). The calculation formula for the GR was (number of germinated seeds/number of seeds tested) $\times 100\%$. ST-GR was defined as the (GR under salt treatment/the GR under control) $\times 100\%$, which represents the ST of soybean accessions.

2.6 | Marker acquisition and candidate gene-based association analysis

The single-nucleotide polymorphism (SNP) markers present in all the 32 MAPKs genes that were used for the candidate gene-based association mapping were obtained from the whole-genome resequencing (WGRS) genotype dataset of 219 soybean accessions (Lu et al., 2020). For this, the SNPs from the promoter region (3 kb before ATG) to the 3'-UTR with minor allele frequency (MAF) $\geq 5\%$ of the 32 MAPKs were extracted. Finally, a total of 1202 SNPs were used as genotypic data for the candidate gene-based association study.

Phenotypic data of ST-GR from the seeds collected from two environments (E1 and E2) were used to perform the association study. The analysis was conducted by TASSEL v5.0 (Bradbury et al., 2007) using a compressed mixed linear model (cMLM; Zhang et al., 2010). For the identification of the significant marker-trait associations, we used two thresholds viz., $P < 0.01$ and $P < 0.001$.

Nucleotide polymorphisms, including SNPs and insertions-deletions (InDels), for the association study of *GmMMK1*, were also identified from the WGRS dataset of 219 soybean accessions (Lu et al., 2020). A total of 101 variants were identified in the *GmMMK1* gene region, and the cMLM method was used to identify significantly marker-trait associations. The Plink 2.0 software (<http://www.cog-genomics.org/plink/2.0/>) was used to calculate the R^2 among the 101 polymorphic sites. Then, the matrix was imported into the LD heatmap package in R v3.6 to draw the linkage disequilibrium (LD) heatmap. Haplotype analysis of *GmMMK1* was performed using the function for clustering genotypes in TASSEL to distinguish haplotypes with polymorphic sites (Bradbury et al., 2007).

2.7 | *Agrobacterium rhizogenes*-mediated transformation of soybean hairy roots

The full-length coding sequence of *GmMMK1* from Williams 82 was cloned into the vector pMDC83 with CaMV 35S promoter (35S::*GmMMK1*). The recombinant pMDC83-*GmMMK1* plasmid vector was transformed into soybean cultivar Williams 82 using *Agrobacterium rhizogenes* strain K599. One-week-old seedlings were used to do soybean cotyledon injection with pMDC83-*GmMMK1* or empty vector and then transferred into a suitable light density, 12/12 h dark/light cycle, day/night temperature of 28/25°C and high (>90%) humidity environment. After almost a month, two kinds of transgenic hairy roots of soybean were developed viz., one overexpressed the *GmMMK1*, and the other one possessed the empty vector control. The detailed procedure used to obtain the transgenic hairy roots was followed as previously described (Kereszt et al., 2007). Plants harboring positive hairy roots with excessive *GmMMK1* were validated using RT-PCR (Figure S8). Moreover, every soybean plant with positive transgenic hairy roots is used for phenotyping and was regarded to be an independent line, which shared different genotypes with others.



2.8 | Phenotype assay of soybean and *Arabidopsis*

When the hairy roots of the transformed soybean started to stretch, we cut the primary roots and transplanted them to hydroponic tanks with 1/2 Hoagland's solution. Four days after growth resumed, tanks were divided into two groups: 1/2 Hoagland's solution with 75 mM NaCl, defined as the treatment group (moderate salt stress), and 1/2 Hoagland's solution alone, defined as the control group (without stress). Ten days later, when the salt-stressed and control plants were at the fourth-node (V4) stage, they were investigated for root and shoot development. The top secondary fully expanded leaves were estimated using a hand-held absorbance-based dual wavelength chlorophyll meter (Konica Minolta SPAD-502). The SPAD-502 chlorophyll meter readings can be used to evaluate the leaf chlorophyll concentration (Uddling et al., 2007). Then, we used the electronic balance to weigh the fresh weight of the shoots and roots, which were separated by cotyledon nodes.

At the same time, the transgenic and control soybeans cultivated in 1/2 Hoagland's for 2 weeks were transferred into 1/2 Hoagland's solution with 150 mM NaCl (severe salt stress). Two days later, transgenic or control soybeans were investigated for ST by checking their live conditions.

For the *Arabidopsis* salt sensitivity assay at the germination stage, seeds from wild-type (Col-0) transformed with empty vector (VC) and transgenic overexpressed *GmMMK1* lines (OE-lines) were surface sterilized, and sown onto 1/2 MS plates containing NaCl at gradient concentrations. The GRs were calculated every 24 h for 7 days, and photographs were taken on the 7th day. For the *Arabidopsis* salt sensitivity assay at the seedling stage, the wild-type and transgenic *Arabidopsis* lines were germinated and grown on 1/2 MS plates for 5 days. Then, the uniformly growing and healthy seedlings were selected and subsequently transferred to plates with 0, 50, or 75 mM NaCl. Photographs were taken on the 5th day after transferring.

2.9 | RNA-seq analysis of transgenic soybean hairy roots

For the RNA-seq analysis, hairy roots in which the green fluorescent protein (GFP) was detected by fluorescence excitation using the hand-held lamp (Luyor Instrument), were collected from plants with transgenic hairy roots and plants with the empty vector control. Total RNA was extracted from the tissue using TRIzol® Reagent according to the manufacturer's instructions (Invitrogen). The detailed procedure for library construction and RNA sequencing was followed as previously described (Sharmin et al., 2020). RNA quantity and integrity were checked using Nano-Drop Spectrometer (Thermo Scientific) and Agilent Bioanalyzer 2100 (Agilent Technologies), respectively. For library construction, an integrity value ≥ 7 was set to select the best quality RNA. A total of 2 μ g RNA from each sample was used for the construction of paired-end libraries using Illumina TruSeq RNA Sample Preparation Kit (Illumina), in accordance with the manufacturer's instructions. Qubit® 2.0 Fluorometer (Life Technologies) was used to quantify purified libraries, and are authenticated by Agilent 2100-bioanalyzer (Agilent Technologies) to validate the

fragment size and concentration. Furthermore, libraries were diluted up to 10 pmol for cBot clustering, and RNA-Seq was performed using Illumina HiSeq 2500 platform by Shanghai Biotechnology Corporation. Short-fragment reads, rRNA reads, adapter/primer, and low-quality reads were filtered from Fastq files of raw data by using Trimmomatic v 0.35 (Bolger et al., 2014) and FASTX-toolkit (Gordon & Hannon, 2010). High-quality filtered reads were mapped to the soybean reference genome (*G. max*1.1 version; Goodstein et al., 2012; Schmutz et al., 2010) using the HISAT2 software (Kim et al., 2015). Reads with unique mapping were used for further studies. Transcriptional profiles (FPKM) of all the samples and differentially expressed genes (DEGs) between the salt stress and control soybean plants were measured using Cufflinks software (Trapnell et al., 2012). Transcripts with a $\log_2 FC$ (fold change) $\geq +2$ and ≤ -2 , and FDR (false discovery rate) value < 0.05 were selected as significant DEGs.

To understand the functions of the DEGs, Gene Ontology (GO) functional enrichment and KEGG pathway analysis were carried out by goatools (<https://github.com/tanghaibao/Goatools>) and KOBAS (<http://kobas.cbi.pku.edu.cn/home.do>), respectively. DEGs were significantly enriched in GO terms and metabolic pathways when their Bonferroni-corrected P-value was less than 0.05.

2.10 | Yeast two-hybrid (Y2H) library screening

A soybean root cDNA library was generated by following the manufacturer's instructions (Clontech). The full-length coding sequence of *GmMMK1* gene was fused to the GAL4 DNA-binding domain of pGBK7 vector (bait vector), and the recombinant vector was transformed into the yeast strain named Y2H Gold. Yeast two-hybrid (Y2H) library screening was performed by using the BD Matchmaker Library Construction and Screening Kit (Clontech). Then, to confirm the interaction obtained by Y2H screening, the *GmMMK1* interacting proteins were subcloned into the pGADT7 (prey vector). Subsequently, the prey vector was co-transformed with the bait vector into Y2H Gold. The pGBK7-53 and pGBK7-Lam bait plasmids were co-transformed with the pGADT7-T prey plasmid as positive and negative controls, respectively.

3 | RESULTS

3.1 | Genome-wide characterization and phylogenetic analysis of the MAPKs in soybean

In total, the reanalysis of the MAPK gene family in soybean by using the updated version of soybean reference genome sequence (soybean genome v4.0) has identified the 32 GmMAPKs in the whole soybean genome (Table S1). These newly confirmed 32 MAPKs were randomly distributed among the 15 of the total 20 soybean chromosomes. Protein length (i.e., number of amino acids) of the GmMAPKs predicted varied considerably from 279 (*Glyma.11G021800*) to 615 (*Glyma.18G106800*) amino acids. Similar to *Arabidopsis* (Jonak et al., 2002), the soybean

MAPKs also possessed the conserved phosphorylated amino acid residues viz., threonine (T) and tyrosine (Y) located on the activation loop of the kinases (Figures S1–S3).

To further elucidate the subfamily grouping and their functional classification, the phylogenetic analysis was performed using MAPKs from *Glycine Max* (*G. max*), *Arabidopsis thaliana*, and *O. sativa*. A neighbor-joining (NJ) phylogenetic tree was constructed that classified the MAPKs into four major groups viz., A, B, C, and D (Figure 1). Each group contains MAPKs from three different plant species viz., *G. max*, *A. thaliana*, and *O. sativa*. However, most of the soybean MAPKs were clustered more closely with the *Arabidopsis* MAPKs compared to the *O. sativa*. Interestingly, soybean had a much higher proportion (31.3%, 10 members) of

GmMAPKs in group B compared to *Arabidopsis* (25.0%, 5 members) and *O. sativa* (11.8%, 2 members). Among the 10 members of soybean group B MAPKs, the homolog of *AtMPK4* occupies almost half of them (four members). According to the results of collinearity analysis, it revealed the collinearity relationship of these four genes (Figure S4).

3.2 | Expression patterns of MAPKs under different environmental stresses

In the present study, our objective was to determine the role of the *GmMAPKs* against various environmental stresses. In this regard, we

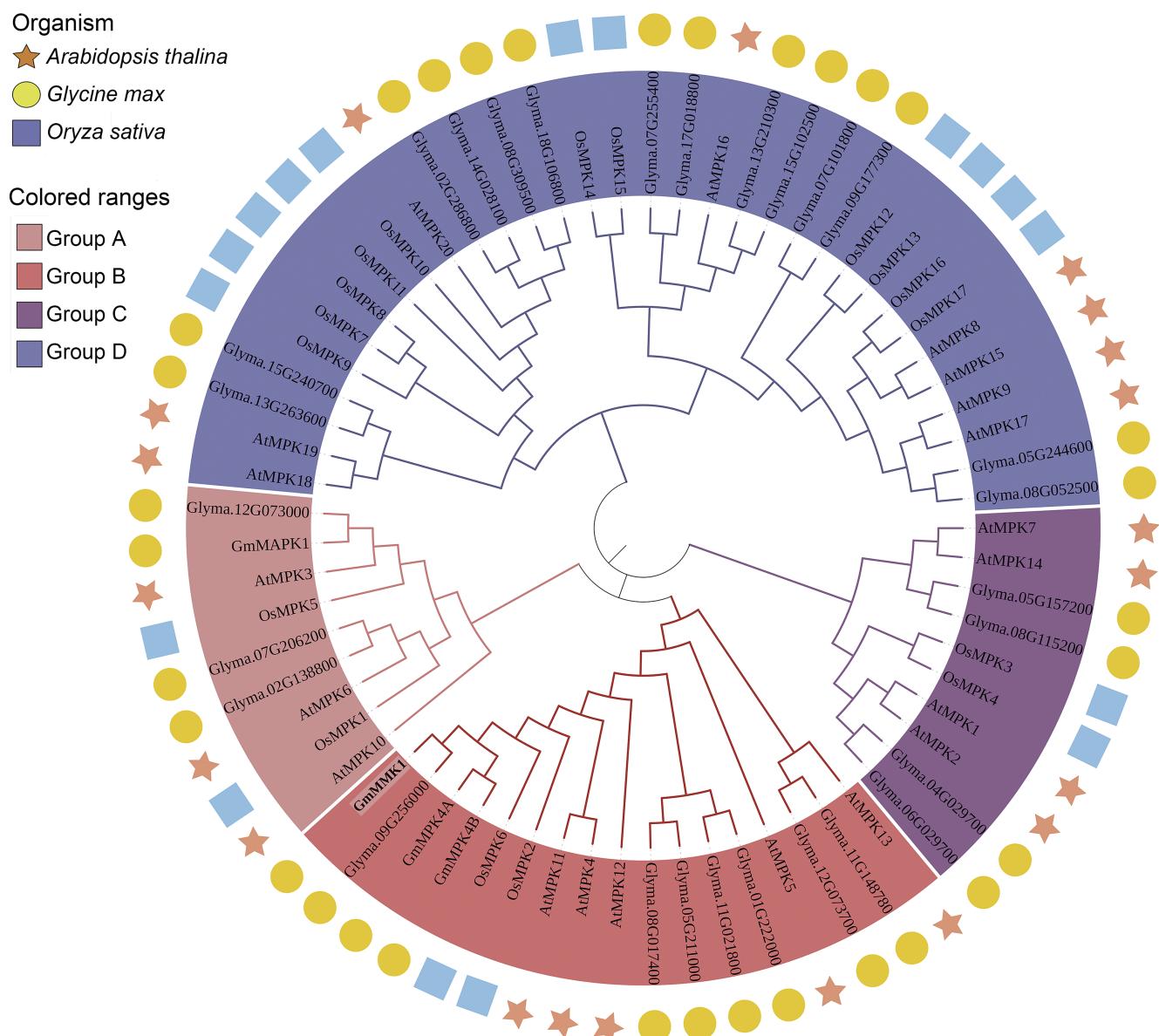


FIGURE 1 Phylogenetic relationship of putative MAPKs in *Arabidopsis thaliana*, *Oryza Sativa*, and *Glycine Max*. The phylogenetic tree was created using the NJ (Neighbor-Joining) method. Bootstrap values were determined by 1000 replicates and are indicated on each branch. Four colors indicate four subgroups viz., A, B, C, and D groups of MAPKs, and three kinds of shapes represent two dicots viz., *Arabidopsis* and soybean; one monocot viz., *O. Sativa*

collected nine RNA-seq or microarray datasets (Table S2), which are freely available online. They reflected soybean gene expression under different environmental stresses viz., 150 mM NaCl treatment, 100 mM NaCl treatment, 50 mM NaHCO₃ treatment, dehydration treatment, aluminum treatment, *Spodoptera litura* (*Fabricius*) feeding, soybean aphid infestation, *Phytophthora sojae* infection, and rust disease infection (Figure 2A,B). Under abiotic stress treatments, the expressions of 21 MAPKs were observed to show the significant change (fold change >1.5), while only four members revealed significant fold change under biotic stress treatment. Among these stress-responsive genes, the GmMPK4B and GmMMK1 (*Glyma18G236800*) showed significant response to NaCl, NaCHO₃, and aluminum stress treatments (Figure 2B).

Based on the expression results, the salt stress was observed to be the most prominent stress leading to considerable change in the

expression of GmMAPKs. Therefore, the expression patterns of GmMAPKs under salt stress were analyzed in detail (Figure 2C). The results showed that MAPKs, acting as kinases, respond very rapidly to the stimulus like salt stress or osmotic stress. In addition, the basic mRNA abundance was shown in the heatmap (Figure 2D), which revealed a huge difference in the basic expression of MAPKs. Moreover, tissue expression analysis based on the RNA-seq dataset from the SoyBase database showed that the roots and nodules had the highest expression tissue of MAPKs (Figure S5). Through public data analysis, we constructed a basic pattern of the stress-induced expression and tissue expression of MAPKs.

To get a better understanding of salt-related MAPKs, we confirmed the expression pattern of a multiple responsive MAPK member, GmMMK1, as mentioned and discussed above. Using a salt-sensitive soybean accession NJAU_C002, and a salt-resistant

Colored ranges

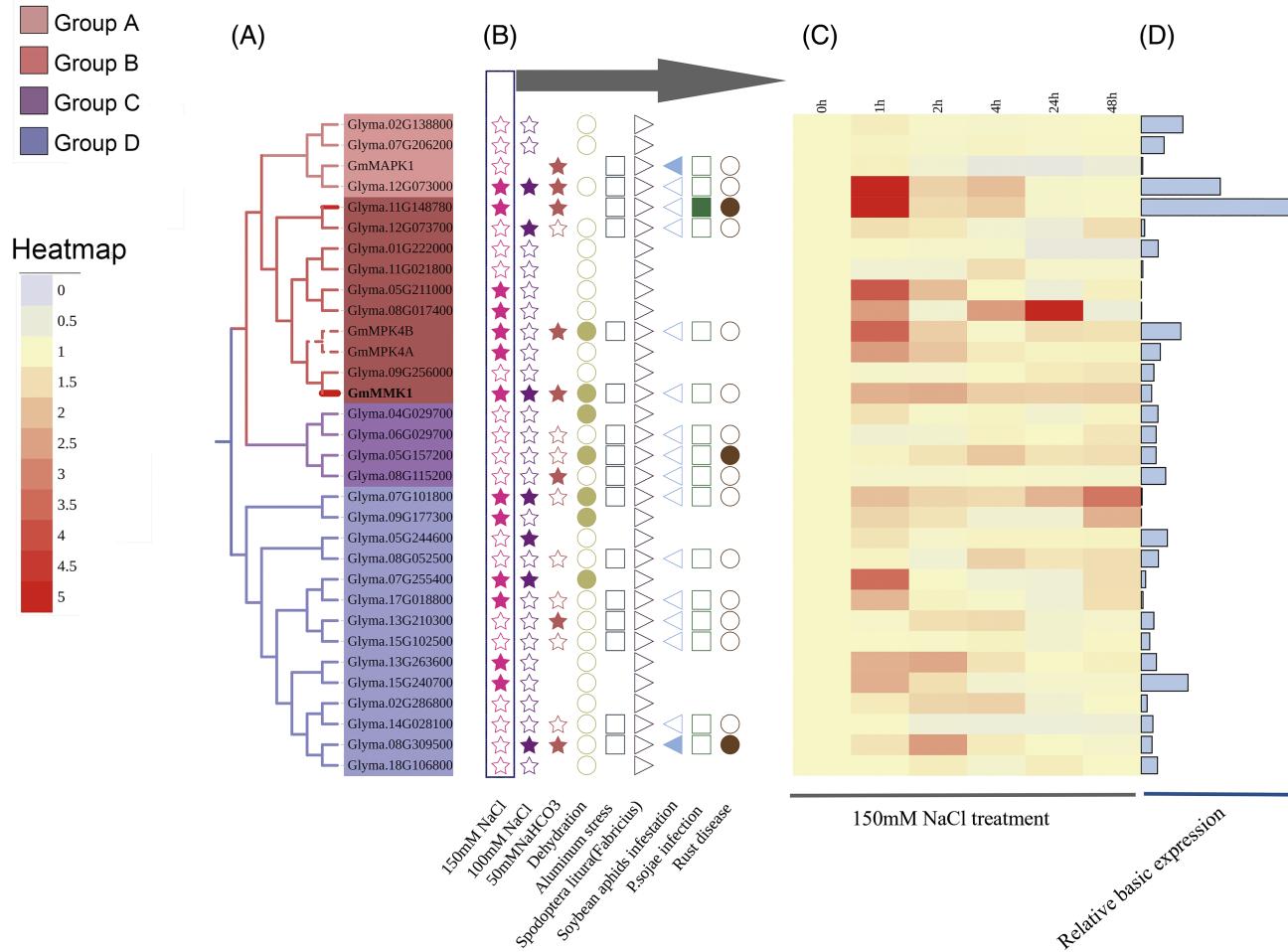


FIGURE 2 Expression patterns of MAPKs under diverse environmental stress treatments. (A) Phylogenetic relationship of putative MAPKs in soybean. (B) The pink star, purple star, red star, gray circle, right triangle, left triangle, green square, and brown circle represent 150 mM NaCl stress, 100 mM NaCl stress, 50 mM NaHCO₃ stress, dehydration stress, aluminum stress, *Spodoptera litura* biting, soybean aphid infestation, *Phytophthora sojae* infection and rust disease, respectively. The blank shapes mean genes in expression profiles were missing, whereas solid shapes represent a significant response to the stress, and hollow shapes represent not responding. The threshold of significant upregulation or downregulation was set as 1.5-fold. These public datasets can be found by the NCBI accession numbers provided in Table S2. (C) The time-course expression profile of soybean root in response to salt stress. (D) Basic mRNA abundance of GmMMK1 in RNA-seq data without salt stress

accession, NJAU_C101 (Figure S6C), we found that NJAU_C002, which is sensitive to salt at the germination stage, had a higher and faster response to salt stress. However, the performance of the NJAU_C101 after salt treatment was exactly the opposite (Figure S6A). Additionally, the expression pattern at the seedling stage was consistent but more evident relative to that at the germination stage (Figure S6B). In salt-sensitive samples, there was a rapid increase of *GmMMK1* and subsequently declined, then increased again in a later period of 48 h. In salt-tolerant samples, the *GmMMK1* seems to be dull, and the expression level was relatively reduced (except 24 h in the seedling stage). More interestingly, at each time point, the expression patterns of sensitive and tolerant materials are reversed.

3.3 | Candidate gene-based association mapping of MAPKs and haplotype analysis of *GmMMK1*

GR of the 219 soybean accessions was evaluated under salt stress for two independent years. Consistent with previous research (Kan et al., 2015), frequency distributions were approximately consistent with a normal distribution (Figure S7). By using the public WGRS project of 424 soybean accessions (Lu et al., 2020), we extracted the high-quality SNPs markers with MAF \geq 5% in the MAPK gene regions (from 3 kb before ATG to the 3'UTR). We identified a total of 1202 high-quality polymorphic SNPs within the 31 *GmMAPKs* except in *Glyma.08G017400* in which no SNP had been detected. An average of 38.8 SNPs was found in each *GmMAPK* gene, with the 97 SNPs located in the *GmMMK1*. The candidate gene-based association study showed that *Glyma.02G138800*, *Glyma.02G286800*, *Glyma.07G101800*, *Glyma.08G052500*, *Glyma.14G028100*, *Glyma.18G106800*, and *Glyma.18G236800* were significantly associated with the ST-GR. Among these genes, the *GmMMK1* (*Glyma.18G236800*) showed the highest number of the significant SNPs associated with the ST-GR. Furthermore, the *GmMMK1* was consistently detected to be associated with the ST-GR in both environments, namely, the mean data and BLUP data ($P < 0.05$). Detailed results are presented in Table S3.

To further determine the relationship between genetic variations in *GmMMK1* and ST-GR phenotype, 93 SNPs and 8 InDels were screened from the resequencing data, and we analyzed the association of these 101 polymorphisms (91 SNPs and 8 InDels) again with the BLUP data, and calculated the significance that is, LD of these polymorphic markers (Figure 3A–C). A total of eight polymorphic markers were observed to show significant trait-marker association ($P < 0.01$), which included seven SNPs and one InDel. Through gene structure annotation, we found that most of the significant polymorphic markers (87.5%) were present in the noncoding regions, especially the promoter region (Table S4). Based on these eight polymorphic significant markers, the 219 soybean accessions were classified into 10 haplotypes (Hap). The top four haplotypes accounted for 95.4% of the total accessions. Hap1, Hap2, Hap3, and Hap4 possess 163, 20, 14, and 12 individuals, respectively, and their mean values of ST-GR were 0.735, 0.605, 0.696, and 0.622, respectively (Figure 3E). The ST-GR of Hap1 allele was significantly higher than those of Hap2 ($P = 0.0055$ by t-test) and Hap4 ($P = 0.033$ by t-test). Further

investigation of the sequence variation of these top four haplotypes alleles showed that INDEL_52570745 of high ST-GR Haps (Hap1 and Hap3) was A, while low ST-GR Haps (Hap2 and Hap4) were missing (Figure 3D). Besides, it is interesting that the two accessions we chose for the expression pattern of *GmMMK1* shared different haplotypes, which NJAU_C101 was Hap1 and NJAU_C002 was Hap2 (Figure S6).

3.4 | *GmMMK1* as a negative regulator of ST in soybean

Results of the bioinformatics, transcriptomic, and association mapping revealed the important role of the *GmMMK1* in the salt stress response, and this gene was a soybean *MPK4* homolog. Hence, we selected the *GmMMK1* gene to validate its function in response to salt stress by using the overexpression analysis. The *GmMMK1*-OE lines represent the transgenic plants overexpressing the *GmMMK1*, whereas the control plants possess the empty vector devoid of *GmMMK1*. Expression analysis showed that *GmMMK1*-OE lines exhibited significantly higher mRNA levels of *GmMMK1* compared to nontransgenic control plants (Figure S8A).

After 10 days of treatment, no significant phenotypic difference was observed between the transgenic and control plants in the presence of 0 mM NaCl. In contrast, under salt stress conditions, clear-cut phenotypic differences were observed between the *GmMMK1*-OE and control plants (Figure 4A). For example, the *GmMMK1*-OE plants exhibited higher chlorosis and damage to the leaves with lower average leaf chlorophyll levels (using SPAD-502 chlorophyll meter values) than the control plants under salt stress (Figure 4C). In addition, we also investigated two morphological parameters viz., the fresh weight of the aerial part (Figure 4D) and the underground part (Figure 4E). The average fresh weight of the underground part of the control plants was significantly higher than that of the *GmMMK1*-OE plants. These results indicated that the root development of *GmMMK1*-OE plants exhibited higher sensitivity to salt stress than the control plants. In addition to the morphological changes under moderate salt stress, the ST of *GmMMK1*-OE and control plants under severe salt stress was investigated. Our results revealed that exposure to 150 mM NaCl for 2 days led to obvious differences between *GmMMK1*-OE and control plants (Figure 4B). The *GmMMK1*-OE plants presented surprising injury symptoms, such as serious wilting, leaf whitening, and early death. However, the control plants exhibited stronger vitality with fewer albino leaves. In general, plants carrying overexpressing *GmMMK1* at the seedling stage appeared to be more sensitive to salt stress, both in terms of morphological performance and final survival ability.

3.5 | Overexpression of *GmMMK1* decreased ST in *Arabidopsis*

Besides the transgenic soybean hairy roots system, we have also developed the transgenic lines of *Arabidopsis* harboring overexpressed

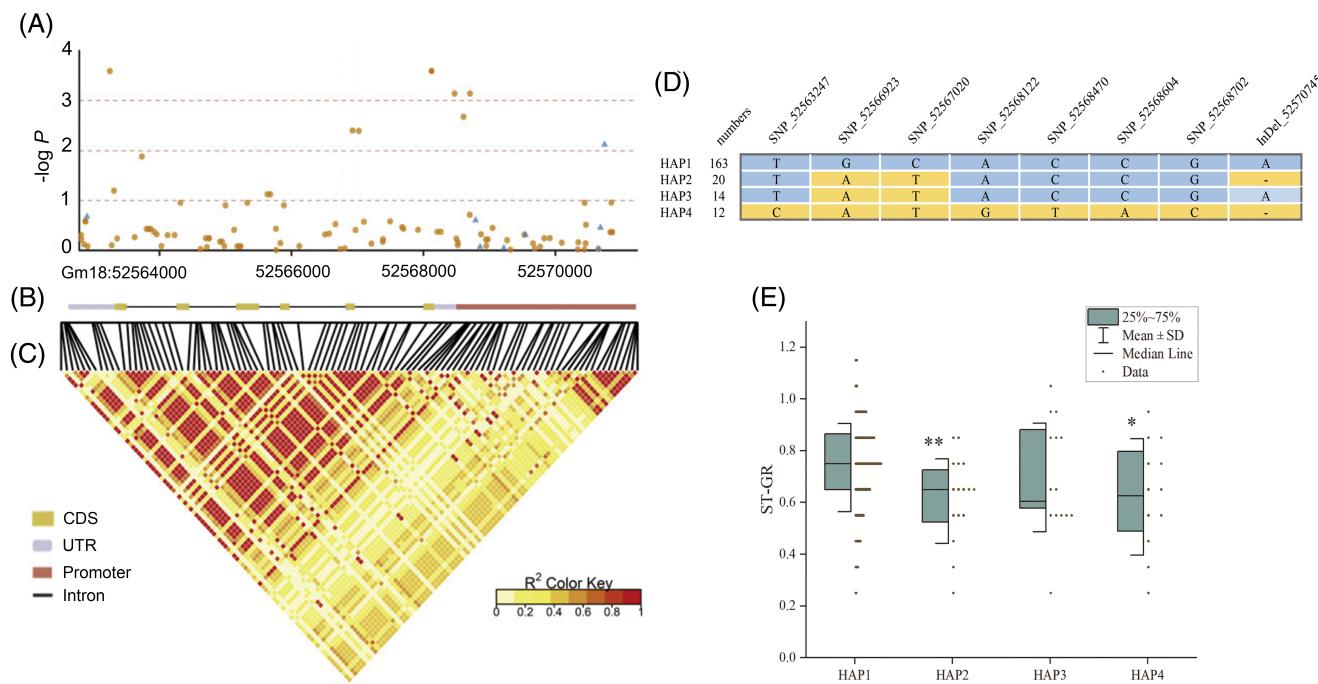


FIGURE 3 Natural variations in GmMMK1 are significantly associated with salt tolerance indices in soybean. (A) Manhattan plot of the association analysis of GmMMK1, including full-length GmMMK1 with the promoter region. The x-axis represents the corresponding physical location on chromosome 18. (B) A schematic diagram of GmMMK1 gene structure. The red rectangle represents the 2-kb-long promoter region, the gray part is the 5'UTR or 3'UTR, the yellow part is the exon, and the black line is the intron of GmMMK1. (C) Linkage disequilibrium (LD) analysis of GmMMK1 gene region. Each small square represents the R^2 value between two sites, red represents complete linkage, and faint yellow represents complete non linkage. (D) Four main haplotypes determined by eight significant SNP/InDels. Yellow represents the type of variation in the first allele frequency and blue represents the type of variation in the minor allele frequency. The symbol “-” stands for that there is a deletion in this haplotype. (E) The ratio of the germination rate under salt conditions to the germination rate under no-salt conditions (ST-GR) of four main haplotypes. The box diagram is divided into two parts: the normal box plot on the left and the data distribution on the right. The ST-GR of Hap1 was tested with Hap2, Hap3, and Hap4 for significance. * $P < 0.05$, ** $P < 0.01$

GmMMK1 to further validate its role in the germination under salt stress. Three T_2 homozygous lines showed higher expression of GmMMK1 than wild-type according to the expression analysis (-Figure S8B). Thus, they were chosen for subsequent experiments, and these lines were designated as OE-11, OE-22, and OE-51. In the absence of salt stress, the control and transgenic lines showed no significant difference in the GR. However, when germinating on the 1/2 MS plates with 50 mM NaCl, it led to a significant decline in the GR of transgenic lines compared to the control plants. Moreover, by increasing the concentration to 75 mM, the GR of the transgenic lines decreased more severely to 50% relative to control plants (Figure 5A). However, when the salt concentration rose to 150 mM, none of the seeds of both transgenic and control plants germinated anymore. Among the three transgenic lines, the OE-11 was highly sensitive to salt stress, while the OE-22, and OE-51 were comparatively less sensitive to salt stress than OE-11, but three lines shared the same tendency to reduce germination (Figure 5B). These results suggested a negative role of GmMMK1 in the regulation of germination in plants.

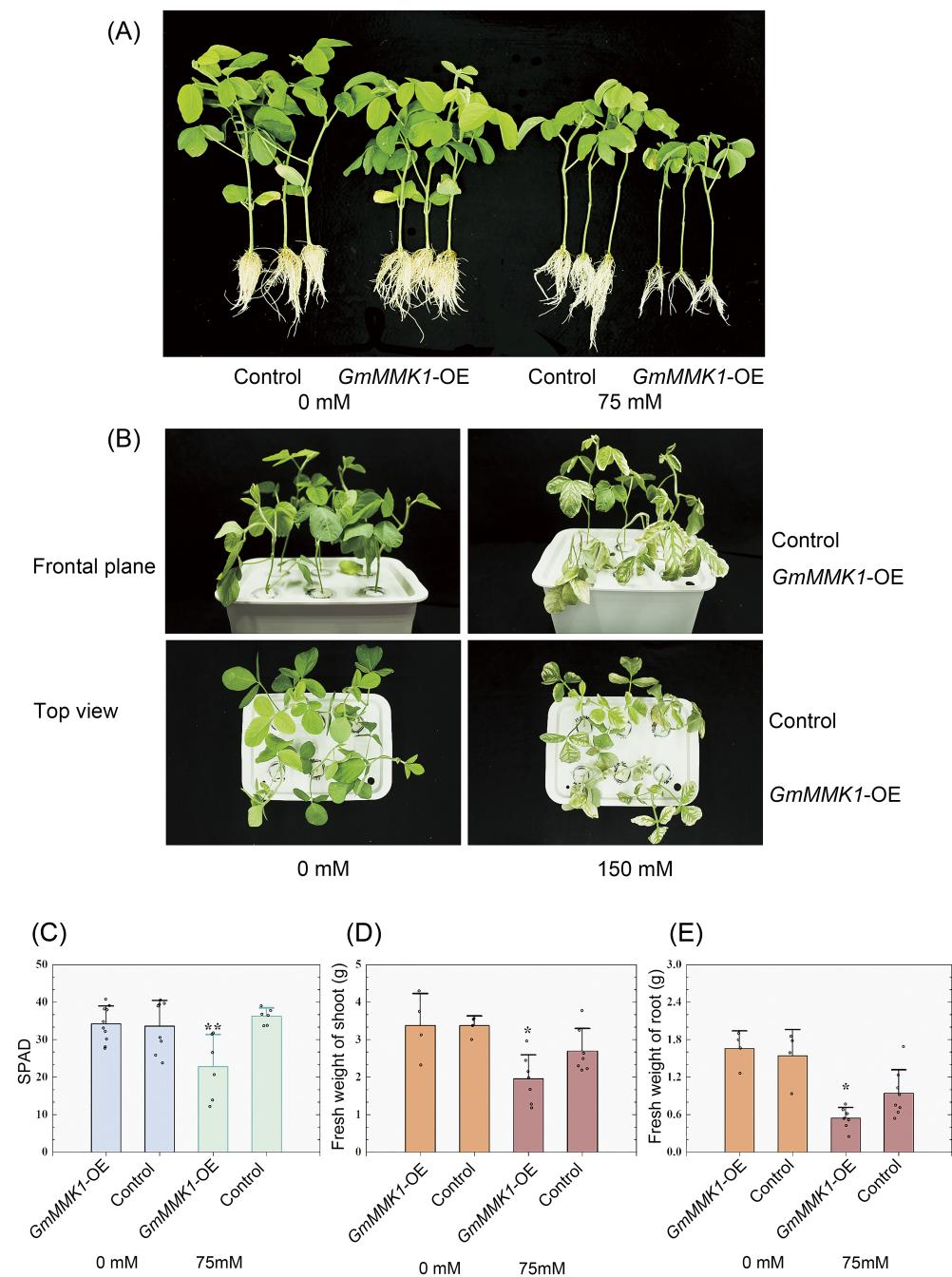
For the ST assay during the seedling stage, the transgenic and wild-type seeds were first germinated, and then transferred to 1/2

MS with 0, 50, and 75 mM NaCl treatment. As shown in Figure 5C,D, the primary root length of transgenic plants was noticeably inhibited under moderate NaCl treatment. The 75 mM NaCl even made *Arabidopsis* seedlings albino and dead, while the control seedlings grew well. These results indicated that overexpression of GmMMK1 in *Arabidopsis* greatly enhanced the seedlings sensitivity to salt stress.

3.6 | Transcriptional regulatory network of GmMMK1 under salt stress

In order to investigate the differences in the transcriptional response of the transgenic and control plants under salt stress, we performed the RNA-seq analysis of GmMMK1-OE soybean hairy roots under salt stress. The expression level of GmMMK1 in the transcriptome of the transgenic lines increased by 22-fold compared to that of control plants, which in turn caused significant up- and down-regulation of 453 and 192 genes, respectively (P -value < 0.05 , fold-change > 2 or < 0.5 , Figure 6A). These DEGs were further subjected to GO enrichment analysis to identify transcriptional changes in transgenic lines in

FIGURE 4 Performance of transgenic soybeans with *GmMMK1*-OE hairy roots under salt stress. (A) Root growth of soybeans with transgenic hairy roots and nontransgenic roots under relatively moderate salt stress (75 mM NaCl). *GmMMK1*-OE represents transgenic soybean chimera with hairy roots overexpressing *GmMMK1*. The control is soybean with hairy roots carrying the empty vector pMDC83. (B) Salt tolerance of soybeans with transgenic hairy roots and nontransgenic roots under severe salt stress (150 mM NaCl). The pictures were taken from the frontal plane and the top. (C–E) (C) SPAD values (chlorophyll values) of the leaves using SPAD-502 meter. (D) Fresh weight of *GmMMK1*-OE and control shoots. (E) Fresh weight of *GmMMK1*-OE and control roots. The data are presented as the mean \pm sds ($n \geq 3$). * $P < 0.05$, ** $P < 0.01$



comparison to control plants. Significant genes were enriched in response to the stimulus (GO: 0050896), with a *P*-value of 1.08×10^{-6} (Figure 6B). Furthermore, 10 genes related to oxygen stress (GO: 0070482, GO: 0036293, and GO: 0001666) were enriched, which suggested that *GmMMK1* may influence the active oxygen scavenging system (Table S5). KEGG analysis showed that in addition to the enrichment of secondary metabolic pathways, the MAPK signaling pathway (KEGG entry: map04016) was also enriched. There were totally 10 genes enriched in the MAPK signaling pathway, while they were not members in the actual MAPK cascade but were possible downstream genes of MAPK signaling such as *PYL4* and *PP2Cs*. *PYL4* is the receptor of ABA, and *PP2Cs* are members of ABA

core signal transduction. In addition, among the 10 genes enriched in the MAPK signaling pathway, nine genes were upregulated, while a disease resistance gene *PR1a* was significantly downregulated (Table S6).

3.7 | Protein–protein interaction network of *GmMMK1*

A subcellular localization experiment revealed that the fluorescence of *GmMMK1* was detected in the nucleus and plasma membrane (Figure 7A), which was consistent with its homologs *GmMAPK4*

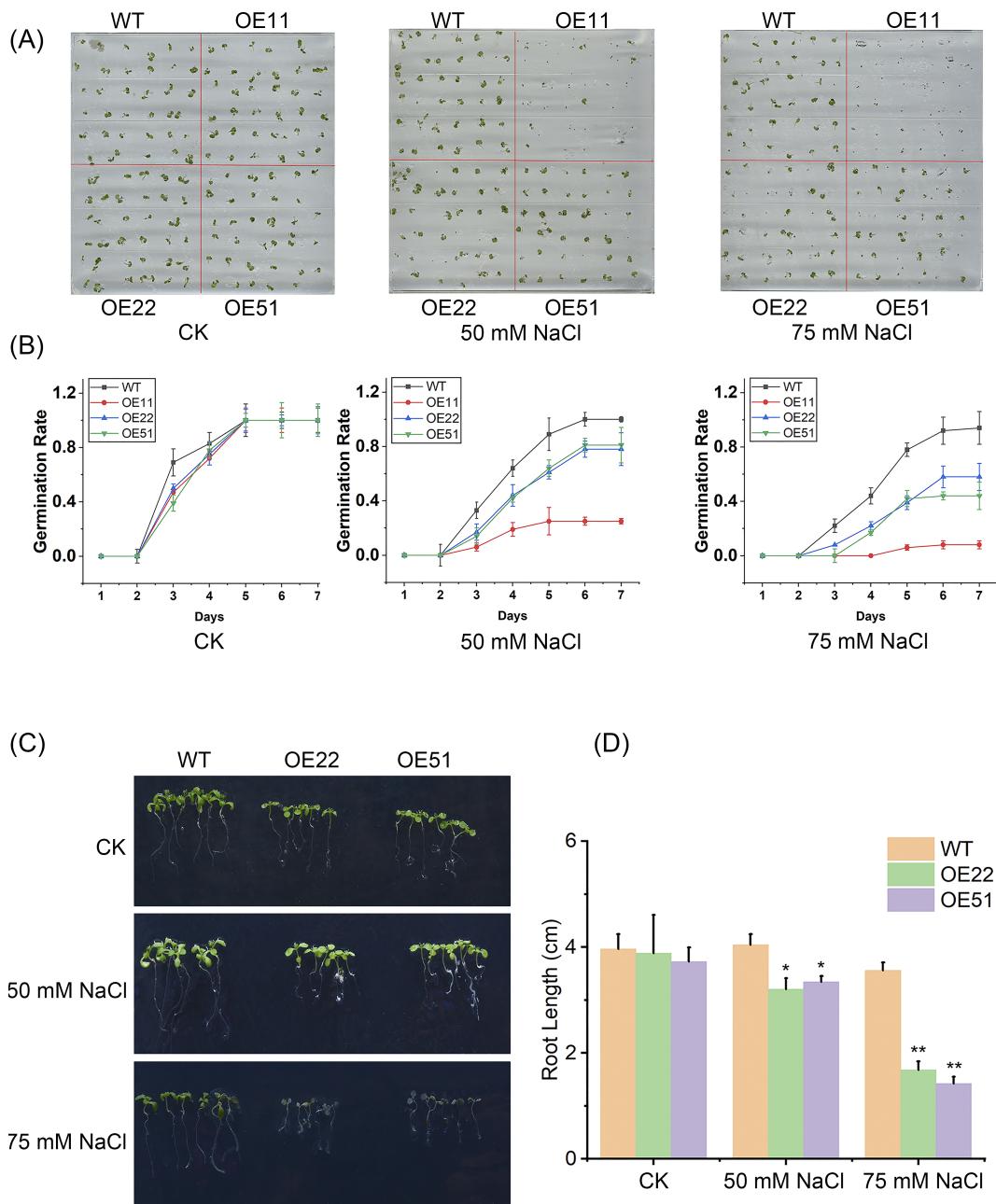


FIGURE 5 Salinity tolerance of transgenic *Arabidopsis* overexpressing *GmMMK1*. (A and B) Germination rate under salt treatment of three NaCl concentration gradients in transgenic *Arabidopsis* homozygous lines and VC (empty vector control lines). The germination rate was calculated every day for a week. OE-21, OE-51, and OE-21 mean three T_2 lines of transgenic *Arabidopsis*. (A) Photographs were taken at the 7th day. (B) Line chart of the daily germination status of each transgenic lines. Data represents the mean \pm SD of three independent replicates. (C and D) Root growth of transgenic and control *Arabidopsis* seedlings. Transgenic *Arabidopsis* and Col-0 were germinated on 1/2 MS and then transferred to gradient salt plates. (C) Photographs were taken at the 5th day after transferring. (D) Histogram of the root growth of transgenic *Arabidopsis* under different salt treatments. Data represents the mean \pm SD of three independent replicates

(Liu et al., 2011) and AtMPK4 (Andreasson et al., 2005). Additionally, the nucleus location of *GmMMK1* was confirmed by co-location with the DAPI signal (Figure S9). Hence, we used a nucleus system of yeast cDNA library to screen the interacting proteins of *GmMMK1*. To this end, we identified the 23 potential interacting proteins (Table 1). Among these proteins, we found some proteins related to hormonal pathways. For example, the *GmSAM1* (Glyma.17G039100) and *GmSAM2*

(Glyma.19G220200) participates in the *S*-adenosylmethionine (SAM) synthesis. SAMs are involved in the process that affects the level of methylation, and they are important precursors of ethylene synthesis. Besides, the *GmPYL4* (Glyma.11G233300) encodes an ABA receptor protein that plays the core ABA signaling role, and its homologous genes were found significantly regulated in the *GmMMK1*-OE transcriptome (Table S6). The *GmOPR* (Glyma.13G109700) is annotated as

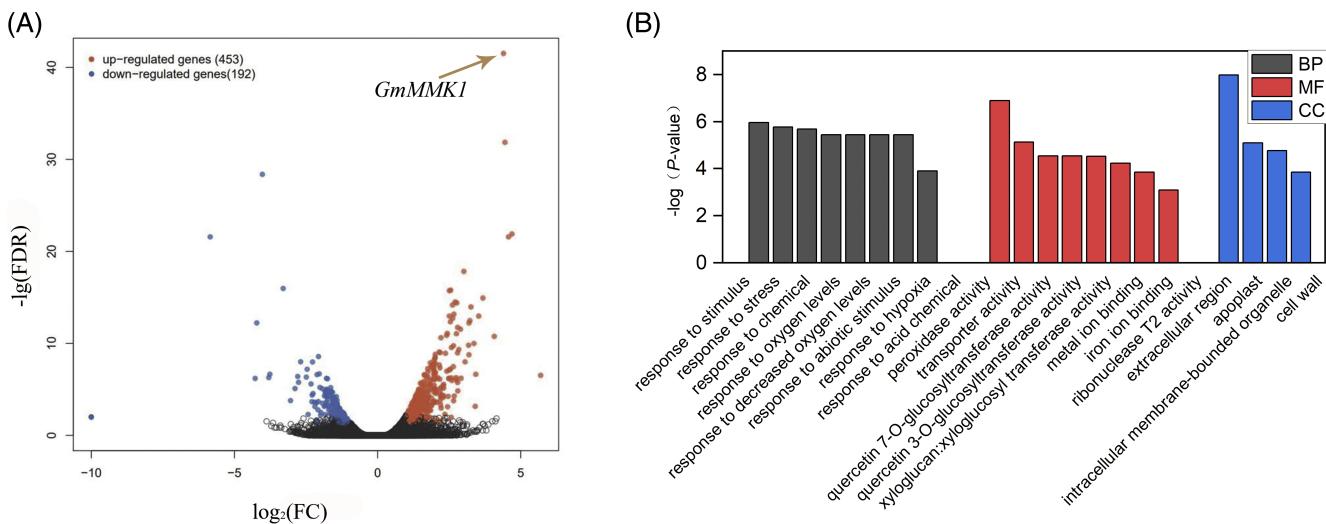


FIGURE 6 Transcriptomic analysis of GmMMK1-OE hairy roots under salt treatment. (A) The x-axis of the volcano map is the log₂ conversion of the fold change (FC). The y-axis represents the log conversion of the false discovery rate (FDR). Orange points indicate upregulated genes, and blue points indicate downregulated genes. (B) Gene Ontology (GO) terms, which are divided into three parts: biological process, molecular function, and cellular component. The x-axis shows the descriptions of significant GO terms. The y-axis shows the log conversion of the P-value

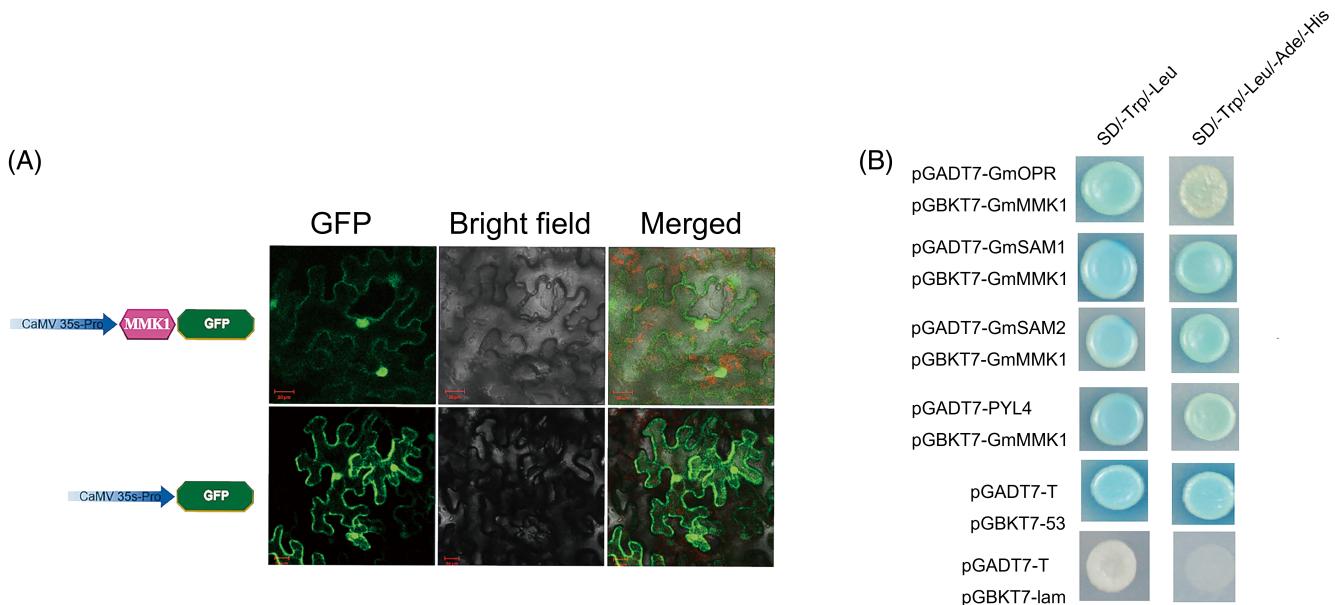


FIGURE 7 GmMMK1 interacting proteins screening by yeast two-hybrid. (A) Subcellular localization of the GmMMK1 protein in the leaves of *Nicotiana benthamiana*. A construct with GmMMK1 fused to a fluorescent protein (GmMMK1-GFP) was generated. The vector control was also transformed into *Agrobacterium tumefaciens* EHA105 and infiltrated into 4-week-old *Nicotiana benthamiana* plants. Scale bars = 20 μ m. (B) Interactions of GmMMK1 with GmOPR, GmSAM1, GmSAM2, and GmPYL4 in yeast. GmMMK1 was expressed in yeast as a GAL4 activation domain (BD) fusion, and GmPYL4 and GmSAM1 were expressed in yeast as a GAL4 activation domain (AD) fusion. Equal amounts of yeast clones were plated on SD-Leu-Trp + X- α -gal and SD-Leu-Trp-His-Ade + X- α -gal selection plates. pGADT7-T and pGBKT7-53 were used as the negative control, while pGADT7-T and pGBKT7-lam were used as the positive control

12-oxophytodienoate reductase, which encodes the enzyme related to the synthesis of jasmonic acid (JA). From the 23 proteins screened by Y2H library, we picked four interesting hormone-related proteins viz., GmOPR, GmSAM1, GmSAM2, and GmPYL4 to confirm the interaction with GmMMK1 (Figure 7B). In order to find the consensus of proteins obtained by Y2H technique, we performed the gene

enrichment on DAVID, as mentioned above. We got some significant terms such as one-carbon metabolic process (P -value = 0.020) and SAM biosynthetic process (P -value = 0.006; Table S7). These findings indicated that GmMMK1 may interact with not only one specific member, but also multiple similar substrates in the SAM synthesis pathway.

| No. | Locus tag | Annotation | Hits |
|-----|-----------------|--|------|
| 1 | Glyma.02G091900 | Actin | 1 |
| 2 | Glyma.02G132100 | Nitrate reductase | 1 |
| 3 | Glyma.04G255800 | Glucuronoxylan 4-O-methyltransferase 1 | 1 |
| 4 | Glyma.05G044000 | Probable pectate lyase 22-like | 2 |
| 5 | Glyma.05G123500 | p34cdc2 protein kinase | 4 |
| 6 | Glyma.06G231000 | UBP1-associated protein 2C | 1 |
| 7 | Glyma.07G132900 | Oxysterol-binding protein-related protein 3A | 1 |
| 8 | Glyma.07G212800 | Ferredoxin-nitrite reductase | 1 |
| 9 | Glyma.08G177900 | Polyadenylate-binding protein 2-like | 1 |
| 10 | Glyma.09G062900 | Beta-galactosidase | 1 |
| 11 | Glyma.09G195000 | N-acylphosphatidylethanolamine synthase | 1 |
| 12 | Glyma.10G179200 | Protein EXORDIUM-like 5 | 1 |
| 13 | Glyma.11G059200 | L-ascorbate oxidase homolog | 1 |
| 14 | Glyma.11G233300 | Abscisic acid receptor PYL4 | 1 |
| 15 | Glyma.13G005500 | Alpha-amylase inhibitor/lipid transfer/seed storage family protein | 2 |
| 16 | Glyma.13G109700 | putative 12-oxophytodienoate reductase | 1 |
| 17 | Glyma.13G268000 | gamma-glutamyl hydrolase 2 | 1 |
| 18 | Glyma.16G012200 | 40S ribosomal protein SA | 1 |
| 19 | Glyma.16G181700 | Uncharacterized LOC100809067 | 1 |
| 20 | Glyma.17G039100 | S-adenosylmethionine synthase-like | 1 |
| 21 | Glyma.17G052100 | WD repeat-containing protein VIP3 | 1 |
| 22 | Glyma.19G109000 | Uncharacterized LOC100783843 | 1 |
| 23 | Glyma.19G220200 | S-adenosylmethionine synthase 1 | 1 |

TABLE 1 Proteins that interact with GmMMK1 identified via., yeast two-hybrid screening

4 | DISCUSSION

4.1 | Polymorphisms of GmMMK1 are associated with ST in soybean

In this study, we performed association analysis and found a salt-related candidate gene, *GmMMK1*. Through further association analysis of *GmMMK1*, we found that seven significant loci were concentrated in the noncoding regions, which may affect the transcription level of *GmMMK1*. Many genes have been mapped using candidate association analysis, and found to have significant polymorphic sites in noncoding regions such as promoters. Through association analysis of *GmCDF1*, it was found that the relative germination rate of Hap2 of *GmCDF1* under salt treatment was significantly higher than that of Hap1, and the difference between Hap1 and Hap2 was mainly due to the deletion of one base at S-671 in the promoter region (Zhang et al., 2019). Similarly, a waterlogging related gene *ZmERE180* was found in maize (Yu et al., 2019). By association analysis of the candidate gene, it was also found that the gene was divided into two haplotypes by the insertion and deletion of InDel59 in the promoter region, and the survival rate of Hap1 under waterlogging was significantly higher than that of Hap2 (Yu et al., 2019). So, it was interesting to do candidate gene association in MAPKs to find a salt-related MAPK in soybean. Finally we found *GmMMK1*, similar to these genes above,

the 87.5% (7/8) significant variation of *GmMMK1* came from the non-coding region, in which there were four loci in the promoter region. Hap2 and Hap4 possessed a deletion at InDel_52570745, and their ST-GR was significantly lower than that of Hap1 and Hap2. This may imply that the variation in the promoter is the main reason for the functional differences in the germination rate of *GmMMK1* under salt treatment in the natural population. Due to this, the sequencing and haplotype typing of *GmMMK1* may be used to identify salt-tolerant germplasm in further crop breeding.

4.2 | Gene expansion of MAPKs and the complicated role of MPK4 subfamily

MAPK represents the last component of the MAP3K-MAP2K-MAPK cascade. To date, the MAPK gene family has been characterized in the multiple plant species at the whole-genome level, such as *Arabidopsis* (Colcombet & Hirt, 2008), *O. sativa* (Liu & Xue, 2007), and *G. max* (Neupane et al., 2013). Based on the genome scans using the whole genome sequence of the *Williams 82 v4.0* genotype, we identified a total of 32 putative members of the soybean MAPK family. There are six previous MAPKs (Neupane et al., 2013) removed from *GmMAPKs*, which have been annotated as other kinase family members now.

In the present study, through candidate gene association analysis, we focused on the MAPKs of groups B, which contains the TEY motif. Group B is closely related to various hormone signals and participates in the response to disease resistance signals (Rodriguez et al., 2010). Our study reported that members of group B MAPKs were more abundant in the soybean genome (10 Group B members) relative to the monocots such as maize (2 Group B members) (Sun et al., 2015) and rice (2 Group B members). Among these genes, the MPK4 orthologs of *Arabidopsis* constitute an important subfamily. The collinearity analysis of the soybean MAPKs showed that Glyma.16G032900 (*GmMAPK4a*) versus Glyma.07G066800 (*GmMAPK4b*) and *GmMAPK4c* (Glyma.09G25600) versus *GmMAPK4d* (*GmMMK1* or Glyma.18G236800) are gene pairs, they are the product of segmental duplication (Figure S5). The replication event of the soybean genome brings more evolutionary possibilities to the function of MAPKs, but the redundancy also brings certain difficulties to the research of MAPKs' function. What's more, the study of homologous MAPKs has no clear answer. For example, the orthologous gene *OsMPK6* in *O. sativa* acts as a downstream effector of *OsSIT1*, and the *SIT1-MPK6* cascade can mediate salt sensitivity by affecting reactive oxygen species (ROS) and ethylene signaling, leading to growth inhibition and plant death under salt stress (Li et al., 2014). However, in *Arabidopsis*, overexpression of *MKK2* upregulated the MPK4 and MPK6 activity, and enhanced ST (Teige et al., 2004). More interestingly, it was reported that *GmMPK6* functioned as both a repressor and an activator in the soybean defense responses (Liu et al., 2014). These multiple functions may result from MAPKs being at the intersection of multiple stress-related signaling. For example, overexpression of *SpMPK3* in *Arabidopsis* enhanced the expression of many ABA-induced genes and affected the germination (Li et al., 2014). Moreover, silencing of *NtMPK4* strongly increased the resistance to two herbivores in a fashion of regulating JA signaling (Hettenhausen et al., 2013). What's more, activation of the endogenous MPK3 and MPK6 kinases led to up-regulation of genes responsible for ethylene biosynthesis and ethylene responses (Xu et al., 2008). In our study, we found *GmMMK1* plays a negative role in ST. Ectopic expression of *GmMMK1* in *Arabidopsis* has led to a significant decrease in the GR of transgenic seeds under salt stress. To determine whether *GmMMK1* is involved in multiple hormone signaling, and how it functions in salt stress signaling, identification of the genetic network of *GmMMK1* is crucial.

4.3 | Negative roles of *GmMMK1* in soybean salt signaling network

Since our objective was to figure out the regulation network of *GmMMK1*, it was easy for us to speculate its upstreaming regulator because of its role in the MAPK cascade. In fact, upstream members of MAPK cascade were identified as early components of salt and osmotic signaling (Fàbregas et al., 2020). Recently, the novel roles for Raf-like kinases (MAPKKK members) in osmotic stress pathways have been identified as regulators of ABA-dependent and ABA-independent signaling pathways. So, the related homologous Raf-like *B3* and *B4* family genes expression from the time-course salt

treatment transcriptome were analyzed. We found that soybean *B4* and *B3 Raf* families also have a significant response to salt stress within a very short time, which indicates that the expression of MAPKs might be regulated by their upstream members (Figure S10). In the present study, the response to salt stress of the MAPKs at transcription level had been well demonstrated (Figure 2). Through our and previous evidence, we believed that the whole MAPK cascade plays an early component in salt signaling. However, different MAPKs have different functions, and it may be because of their specific substrate. Hence, to understand the downstream regulatory network of *GmMMK1* related to ST at the transcriptional level, we carried out the RNA-seq analysis of *GmMMK1-OE* hairy roots. Our results reported that oxidative stress-related genes (GO: 0006979) were enriched among the downregulated genes (Table S5), which indicated that *GmMMK1* may reduce the active oxygen scavenging ability of soybean. Salinity-induced ROS accumulation also acts as a signal in the mediation of salinity tolerance (Mittler, 2017), while the ROS homeostasis in transgenic hairy roots has been changed and this might bring negative effects on *GmMMK1-OE* plants. At the same time, the ABA core signaling members namely, PYLs and PP2Cs exhibited a significant increase in the transcriptional levels (Table S6). Overexpression of *GhPYL10/12/26* in the *Arabidopsis* has led to increased sensitivity of transgenic plants for the seed germination and early seedling growth to ABA (Chen et al., 2017). Similarly, *GmPYL4-1* (Glyma.02G261900) and *GmPYL4-2* (Glyma.14G056300) increase expression by twice in *GmMMK1-OE* hairy roots. Interestingly, another *GmPYL4* (Glyma.11G233300) was obtained by Y2H screening of *GmMMK1*. Furthermore, the other ABA-dependent salt stress signaling pathway negative regulator, PP2Cs, also significantly affected by overexpression of *GmMMK1*. Three PP2Cs (Glyma.14G162100, Glyma.01G225100, and Glyma.11G018000) increase 6.3 times in expression on average (Table S6). Among them, Glyma.11G018000 increased 10.3 times with a *P*-value of 9.40^{E-14} , which was annotated as *probable protein phosphatase 2C 8*. Studies showed that overexpression of PP2Cs in *Arabidopsis* led to more sensitivity than the wild-type under 100 or 150 mM NaCl treatment (Chu et al., 2021). So, the out of control of ABA signaling may be another key reason for the negative effect when overexpressing *GmMMK1*.

ACKNOWLEDGMENTS

The authors thank Dr. Dezhou Hu and Dr. Xiao Li at Nanjing Agricultural University for reviewing the language in the manuscript. The authors thank Mr. Qing Du and Mr. Rui Qin for the soybean hairy roots generation experiment. This work was supported in part by the Natural Science Foundation of Jiangsu Province (BK20191313), Ministry of Science and Technology (2017YFE0111000), and EUCLEG Horizon 2020 of European Union (727312).

AUTHOR CONTRIBUTIONS

Guizhen Kan, Xianzhong Feng, and Deyue Yu designed this research; Xiliang Liao, Meiqi Shi, Yali Li, and Wei Zhang performed experiments; Xiliang Liao and Qian Ye performed transcriptome collection and analysis; Xiliang Liao and Meiqi Shi wrote this manuscript. Javaid Akhter Bhat revised the manuscript. All authors approved the manuscript.



DATA AVAILABILITY STATEMENT

The data that support the finding of this study are openly available at the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>), the NCBI (<https://www.ncbi.nlm.nih.gov/>), the TAIR (<https://www.arabidopsis.org/>), the UniProt (<https://www.uniprot.org/>), the SoyBase database (<https://www.soybase.org/soyseq/>) and the plant expression database PLEXdb (<https://www.plexdb.org/>). The data that supports the findings of this study are available in the supplementary material of this article. The materials that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

- Xiliang Liao <https://orcid.org/0000-0003-0726-8571>
 Javaid Akhter Bhat <https://orcid.org/0000-0001-9136-6275>
 Guizhen Kan <https://orcid.org/0000-0003-3252-6106>

REFERENCES

- Andreasson, E., Jenkins, T., Brodersen, P., Thorgrimsen, S., Petersen, N.H.T., Zhu, S. et al. (2005) The MAP kinase substrate MKS1 is a regulator of plant defense responses. *The EMBO Journal*, 24, 2579–2589.
- Andreasson, E. & Ellis, B. (2010) Convergence and specificity in the *Arabidopsis* MAPK nexus. *Trends in Plant Science*, 15, 106–113.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L. et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37, W202–W208.
- Bhat, M.A., Kumar, V., Bhat, M.A., Wani, I.A., Dar, F.L., Farooq, I. et al. (2020) Mechanistic insights of the interaction of plant growth-promoting rhizobacteria (PGPR) with plant roots toward enhancing plant productivity by alleviating salinity stress. *Frontiers in Microbiology*, 11, 1952.
- Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. & Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633–2635.
- Chen, Y., Feng, L., Wei, N., Liu, Z.-H., Hu, S. & Li, X.-B. (2017) Overexpression of cotton PYL genes in *Arabidopsis* enhances the transgenic plant tolerance to drought stress. *Plant Physiology and Biochemistry*, 115, 229–238. <https://doi.org/10.1016/j.plaphy.2017.03.023>
- Chu, M., Chen, P., Meng, S., Xu, P. & Lan, W. (2021) The *Arabidopsis* phosphatase PP2C49 negatively regulates salt tolerance through inhibition of AtHKT1;1. *Journal of Integrative Plant Biology*, 63, 528–542.
- Colcombet, J. & Hirt, H. (2008) *Arabidopsis* MAPKs: a complex signalling network involved in multiple biological processes. *The Biochemical Journal*, 413, 217–226.
- Fàbregas, N., Yoshida, T. & Fernie, A.R. (2020) Role of Raf-like kinases in SnRK2 activation and osmotic stress response in plants. *Nature Communications*, 11, 6184.
- Farooq, M., Gogoi, N., Hussain, M., Barthakur, S., Paul, S., Bharadwaj, N. et al. (2017) Effects, tolerance mechanisms and management of salt stress in grain legumes. *Plant Physiology and Biochemistry*, 118, 199–217.
- FERCHA, A., CAPRIOTTI, A.L., CARUSO, G., CAVALIERE, C., STAMPACIACCHIERI, S., ZENEZINI CHIOZZI, R. et al. (2016) Shotgun proteomic analysis of soybean embryonic axes during germination under salt stress. *Proteomics*, 16, 1537–1546.
- GOODSTEIN, D.M., SHU, S., HOWSON, R., NEUPANE, R., HAYES, R.D., FAZO, J. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40, D1178–D1186.
- GORDON A. & HANNON G.J. (2010). Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished). Available at http://hannonlab.cshl.edu/fastx_toolkit, 5.
- HETTENHAUSEN, C., BALDWIN, I.T. & WU, J. (2013) *Nicotiana attenuata* MPK4 suppresses a novel jasmonic acid (JA) signaling-independent defense pathway against the specialist insect *Manduca sexta*, but is not required for the resistance to the generalist *Spodoptera littoralis*. *New Phytologist*, 199(3), 787–799. <https://doi.org/10.1111/nph.12312>
- ISAYENKOV, S.V. & MAATHUIS, F.J.M. (2019) Plant salinity stress: many unanswered questions remain. *Frontiers in Plant Science*, 10, 80.
- JI, W., CONG, R., LI, S., LI, R., QIN, Z., LI, Y. et al. (2016) Comparative proteomic analysis of soybean leaves and roots by iTRAQ provides insights into response mechanisms to short-term salt stress. *Frontiers in Plant Science*, 7, 573.
- Jonak, C., Okrész, L., Bögrel, L. & Hirt, H. (2002) Complexity, cross talk and integration of plant MAP kinase signalling. *Current Opinion in Plant Biology*, 5, 415–424.
- KAN, G., NING, L., LI, Y., HU, Z., ZHANG, W., HE, X. et al. (2016) Identification of novel loci for salt stress at the seed germination stage in soybean. *Breeding Science*, 66, 530–541.
- KAN, G., ZHANG, W., YANG, W., MA, D., ZHANG, D., HAO, D. et al. (2015) Association mapping of soybean seed germination under salt stress. *Molecular Genetics and Genomics*, 290, 2147–2162.
- KERESZT, A., LI, D., INDRASUMUNAR, A., NGUYEN, C.D.T., NONTACHAIYAPOOM, S., KINKEMA, M. et al. (2007) *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nature Protocols*, 2, 948–952.
- KIM, D., LANGMEAD, B. & SALZBERG, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12, 357–360.
- KUMAR, S., STECHER, G. & TAMURA, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33, 1870–1874.
- LI, C., CHANG, P.-P., GHEBREMARIAM, K.M., QIN, L. & LIANG, Y. (2014) Overexpression of tomato *SpMPK3* gene in *Arabidopsis* enhances the osmotic tolerance. *Biochemical and Biophysical Research Communications*, 443, 357–362.
- LIU, A., XIAO, Z., LI, M.-W., WONG, F.-L., YUNG, W.-S., KU, Y.-S. et al. (2019) Transcriptomic reprogramming in soybean seedlings under salt stress. *Plant, Cell & Environment*, 42, 98–114.
- LIU, J.-Z., BRAUN, E., QIU, W.-L., SHI, Y.-F., MARCELINO-GUIMARÃES, F.C., NAVARRE, D. et al. (2014) Positive and negative roles for soybean MPK6 in regulating defense responses. *Molecular Plant-Microbe Interactions*, 27, 824–834.
- LIU, J.-Z., HORSTMAN, H.D., BRAUN, E., GRAHAM, M.A., ZHANG, C., NAVARRE, D. et al. (2011) Soybean homologs of MPK4 negatively regulate defense responses and positively regulate growth and development. *Plant Physiology*, 157, 1363–1378.
- LIU, Q. & XUE, Q. (2007) Computational identification and phylogenetic analysis of the MAPK gene family in *Oryza sativa*. *Plant Physiology and Biochemistry*, 45, 6–14.
- LIU, Y., ZHOU, M., GAO, Z., REN, W., YANG, F., HE, H. et al. (2015) RNA-Seq analysis reveals MAPKKK family members related to drought tolerance in maize. *PLoS One*, 10, e0143128.
- LU, S., DONG, L., FANG, C., LIU, S., KONG, L., CHENG, Q. et al. (2020) Stepwise selection on homeologous PRR genes controlling flowering and maturity during soybean domestication. *Nature Genetics*, 52, 428–436.
- MAPK GROUP. (2002) Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in Plant Science*, 7, 301–308.
- MENG, X. & ZHANG, S. (2013) MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology*, 51, 245–266.
- MITTLER, R. (2017) ROS are good. *Trends in Plant Science*, 22, 11–19.
- NEUPANE, A., NEPAL, M.P., PIYA, S., SUBRAMANIAN, S., ROHILA, J.S., REESE, R.N. et al. (2013) Identification, nomenclature, and evolutionary relationships of mitogen-activated protein kinase (MAPK) genes in soybean. *Evolutionary Bioinformatics Online*, 9, 363–386.
- PHANG, T.-H., SHAO, G. & LAM, H.-M. (2008) Salt tolerance in soybean. *Journal of Integrative Plant Biology*, 50, 1196–1212.
- PITZSCHEKE, A. (2015) Modes of MAPK substrate recognition and control. *Trends in Plant Science*, 20, 49–55.

- Rodriguez, M.C.S., Petersen, M. & Mundy, J. (2010) Mitogen-activated protein kinase signaling in plants. *Annual Review of Plant Biology*, 61, 621–649.
- Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W. et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, 463, 178–183.
- Sharmin, R.A., Bhuiyan, M.R., Lv, W., Yu, Z., Chang, F., Kong, J. et al. (2020) RNA-Seq based transcriptomic analysis revealed genes associated with seed-flooding tolerance in wild soybean (*Glycine soja* Sieb. & Zucc.). *Environmental and Experimental Botany*, 171, 103906.
- Sun, W., Chen, H., Wang, J., Sun, H.W., Yang, S.K., Sang, Y.L. et al. (2015) Expression analysis of genes encoding mitogen-activated protein kinases in maize provides a key link between abiotic stress signaling and plant reproduction. *Functional & Integrative Genomics*, 15, 107–120.
- Teige, M., Scheikl, E., Eulgem, T., Dóczsi, R., Ichimura, K., Shinozaki, K. et al. (2004) The MKK2 pathway mediates cold and salt stress signaling in *Arabidopsis*. *Molecular Cell*, 15, 141–152.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R. et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7, 562–578.
- Uddling, J., Gelang-Alfredsson, J., Piikki, K. & Pleijel, H. (2007) Evaluating the relationship between leaf chlorophyll concentration and SPAD-502 chlorophyll meter readings. *Photosynthesis Research*, 91, 37–46.
- Valliyodan, B., Cannon, S.B., Bayer, P.E., Shu, S., Brown, A.V., Ren, L. et al. (2019) Construction and comparison of three reference-quality genome assemblies for soybean. *The Plant Journal*, 100, 1066–1082.
- Wang, H., Gong, M., Guo, J., Xin, H., Gao, Y., Liu, C. et al. (2018) Genome-wide identification of *Jatropha curcas* MAPK, MAPKK, and MAPKKK gene families and their expression profile under cold stress. *Scientific Reports*, 8, 16163.
- Xu, J., Li, Y., Wang, Y., Liu, H., Lei, L., Yang, H. et al. (2008) Activation of MAPK Kinase 9 induces ethylene and camalexin biosynthesis and enhances sensitivity to salt stress in *Arabidopsis*. *Journal of Biological Chemistry*, 283(40), 26996–27006. <https://doi.org/10.1074/jbc.m801392200>
- Yu, F., Liang, K., Fang, T., Zhao, H., Han, X., Cai, M. et al. (2019) A group VII ethylene response factor gene, *ZmERE80*, coordinates waterlogging tolerance in maize seedlings. *Plant Biotechnology Journal*, 17, 2286–2298.
- Zhang, W., Liao, X., Cui, Y., Ma, W., Zhang, X., Du, H. et al. (2019) A cation diffusion facilitator, *GmCDF1*, negatively regulates salt tolerance in soybean. *PLoS Genetics*, 15, e1007798.
- Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R.J., Tiwari, H.K., Gore, M.A. et al. (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42, 355–360.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Liao, X., Shi, M., Zhang, W., Ye, Q., Li, Y., Feng, X. et al. (2021) Association analysis of *GmMAPKs* and functional characterization of *GmMMK1* to salt stress response in soybean. *Physiologia Plantarum*, 1–15. Available from: <https://doi.org/10.1111/ppl.13549>