RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2024.109

ScienceAsia 50 (6): 2024: ID 2024109: **1–12**

Genome-wide identification of zf-B-Box gene family in *Glycine max* and expression analysis under NaCl stress

Xiangyu Tan, Xuan Yang, Jinjuan Hao, Xue Gang, Zhenlin Wei*

Biological Sciences Department, Dezhou University, Dezhou, Shandong 253023 China

*Corresponding author, e-mail: wzl19741028@163.com

Received 20 Aug 2023, Accepted 11 Oct 2024
Available online 16 Dec 2024

ABSTRACT: The B-box proteins, or BBX proteins, belong to a subgroup of zinc finger transcription factors and have various physiological functions. This study aimed to conduct a comprehensive genome-wide analysis of soybean *BBX* genes in *Glycine max*. A total of 26 GmBBX proteins were identified. These genes contain 1 or 2 B-box domains, 8 of which have the CCT domain. These proteins were classified into 4 subfamilies based on their structural characteristics and evolutionary relationships. Notably, 18 *GmBBX* genes were identified to form 15 duplicated gene pairs via segmental duplication. In addition, a cross-species syntenic analysis revealed 42 syntenic relationships between 19 *GmBBX* genes and 11 *BBX* genes from *Arabidopsis thaliana*. Nevertheless, there was no apparent synteny between *BBX* genes from *G. max* and *Zea mays*. The transcriptomic data showed that the expression levels of *GmBBX* genes, primarily found in the root and leaf tissues, were generally low, in line with the outcomes from the Ka/Ks ratio analysis. The real-time PCR analysis of 4 selected *GmBBX* genes showed a significant upregulation of gene expression, particularly for *GmBBX20*, under NaCl stress. These findings suggest that GmBBX proteins may be crucial in regulating stress tolerance.

KEYWORDS: *Glycine max*, zinc-finger B-box gene, gene duplication, NaCl stress

INTRODUCTION

Transcription factors (TFs) play a pivotal role in the regulation of gene expression in all eukaryotes through specific protein-protein or protein-DNA interactions [1]. Among the diverse array of TFs, the Zinc-finger B-BOX class stands out as a zinc-finger TF that is remarkably conserved across multicellular organisms and certain unicellular eukaryotes [2].

The BBX proteins have 1 or 2 B-box domains and additional domains that vary depending on the species of origin. Plant BBX proteins are accompanied by the CCT domain (CONSTANS, CO-like, and TOC1) near their C-terminus [2]. Consequently, the B-box domain of plant BBX proteins can function by forming heterodimers with other proteins or within the BBX protein family. This highlights the crucial roles of BBX proteins in mediating protein-protein interaction and transcriptional regulation [3]. Furthermore, the role of the extensively preserved CCT domain, comprising approximately 42–43 amino acid residues, has been associated with the regulation of transcription and the transport of nuclear proteins. The presence of the CCT domain introduces an elevated level of intricacy to the functionalities exhibited by plant BBX proteins [2].

The plant BBX proteins have recently garnered significant attention due to their diverse and vital functions [4–6]. In plants, particularly in *A. thaliana*, AtBBX proteins play a role in controlling growth, development, flowering, shading avoidance, and seedling photomorphogenesis as well as managing biotic and abiotic stresses and plant hormonal pathways [2, 4, 7–9].

The expression of *AtBBX18* gene reduced thermotolerance in *A. thaliana* by controlling the expression of multiple heat-stress-responsive genes [10]. In contrast, overexpression of *SsBBX24* in *A. thaliana* enhanced its salt tolerance. Additionally, genes *OsBBX25* and *CpBBX19* have been found to enhance the salt and drought tolerance of *A. thaliana* [11, 12]. *CmBBX24* has been shown to increase the resistance of *Chrysanthemum morifolium* to drought and low-temperature stress. Many *BBX* genes contain potential *cis*-acting stress-responsive elements, which are consistent with these findings.

G. max is a long-standing economically significant crop used in food production and is an essential nutritional supplement source [13]. However, the land in the primary soybean growing region is susceptible to several abiotic stresses, which negatively impact yield [14]. TFs have been proposed to regulate tolerance to different abiotic stresses. However, the specific BBX protein in *G. max* has not yet been identified or characterized, and it remains uncertain if it plays a role in responding to NaCl and drought stress.

MATERIALS AND METHODS

Sequence retrieval and identification of *BBX* genes in *G. max* genomic sequences

The latest version of the genome sequences and GFF3 annotation files of *G. max*, *Vigna radiata*, *Phaseolus vulgaris*, *Z. mays*, *Oryza sativa* ssp. *Japonica*, *A. thaliana*, and *Vigna angularis* were downloaded from the EnsemblPlant database. The Hidden Markov Model (HMM) profiles of the Zinc-finger B-Box domain (PF00643) were obtained from the PFAM database.

Initially, the entire protein sequences were ob-

tained from the genome sequence, and the B-box domain-containing proteins were identified using the hmmsearch program in HMMER 3.2 with a cut-off *E*-value of $\leqslant 10^{-5}$ [5]. Then, these BBX candidates were further verified by querying the NCBI CDD and SMART databases using default settings [15]. Finally, the remaining proteins were classified as GmBBX proteins based on their chromosomal localization.

The subcellular location of GmBBX proteins was determined by online software WoLF PSORT II (https://www.genscript.com/wolf-psort.html) and Pub-mSubP (http://bioinfo.usu.edu/Plant-mSubP). The basic physicochemical properties of GmBBX proteins, including the theoretical isoelectric point (*pI*), the number of amino acids (aa), the molecular weight (MW), the instability index, and the Grand Average of Hydropathy (GRAVY), were computed using the online server ExPASy.

Gene structure and conserved motif analysis of GmBBXs

The gene structure information (exon/intron organization) of *GmBBX* genes was extracted from its gff3 file and plotted using TBtools software [16]. Furthermore, the conserved motifs were determined by querying MEME server with the following parameters: the number of repetitions, any; the maximum number of motifs, 20; the optimal width of each motif, between 6 and 100 residues [17].

Sequence alignment and phylogenetic analysis of GmBBX proteins

The BBX proteins of *A. thaliana*, *P. vulgaris*, *V. unguiculata*, *V. angularis*, *O. sativa*, and *Z. mays* were identified using the same method as *G. max*. The multiple sequence alignment (MSA) was then performed using Clustal Omega software using full-length protein sequences and B-box domain sequences. Afterwards, the unrooted phylogenetic tree was constructed using the neighbor-joining method of MEGA 7.0 (Possion mode, complete deletion, and 1000 bootstrap values) and was illustrated by Interactive Tree of Life (iTOL) software (https://itol.embl.de/).

Synteny and gene duplication analysis

The multiple collinearity scan toolkit (MCScanX) was applied to identify gene duplication events with default parameters. The orthologous BBX proteins obtained from *G. max*, *Z. mays*, and *A. thaliana* were subjected to syntenic analysis.

In addition, the non-synonymous (Ka) and synonymous (Ks) substitution of each duplicated GmBBX orthologous gene was calculated using KaKs_Calculator 2.0 software [18]. Based on a rate of 6.1×10^{-9} substitutions per site per year, the divergence time (T) was determined as follows: $T = Ks/(2 \times 6.1 \times 10^{-9}) \times 10^{-6}$ Mya [19].

Analysis of promoter regions of *GmBBX* genes

In order to conduct a comprehensive examination of the *cis*-regulatory elements, the promoter regions of all *GmBBX* genes were extracted, specifically targeting the upstream regions spanning 2,000 base pairs. Subsequently, the PlantCARE server was employed to identify and detect the various *cis*-elements present within these promoter regions [20].

The *GmBBX* gene expression analysis using public transcriptomic data

For revealing the expression of *GmBBX* genes in different organs and developmental stages, the transcriptional data were downloaded from the Phytozome V13 database (https://phytozome-next.jgi.doe.gov/), and the corresponding heatmap was displayed using TBtools software [16].

In addition, to identify the *GmBBXs* responsive to abiotic stress, another transcriptional analysis was performed. First, the high-throughput RNA-seq data in Fastq format was retrieved from the Sequence Read Archive (SRA) database with project number PRJNA246058. This dataset comprises the RNA-seq data from leaves treated with salt and drought stresses at various times [21]. After sequence clarification, the expression level of all *G. max* genes was evaluated using Kallisto software as transcripts per kilobase of exon model per million mapped reads (TPM) [22]. Finally, the limma-voom module in the DEApp software was used to identify the differential expressed *GmBBX* genes with FDR-adjusted *p*-values $\leqslant 0.05$ and fold change cutoff of $\geqslant 1.5$ [23].

Quantitative PCR analysis

Uniform and full whole soybean seeds were subjected to surface disinfection with 0.1% mercuric chloride and then germinated in the dark at 25 °C for 48 h. After germination, the NaCl solution in 1/2 Hoagland solution was added to final 150 mM concentrations with 50 mM intervals [24]. This treatment was repeated daily until the true leaves grew. At the same time, seedlings were treated with 1/2 Hoagland solution alone as a control treatment.

From each sample, the RNAs were isolated using an EASYspin Plus Plant RNA Kit (RN56, Aidlab, Beijing, China) from 50 mg of leaves. The RNA quality was analyzed using the Nanodrop 2000 analyzer. The cDNAs were obtained using a PrimeScript™ RT reagent kit with gDNA Eraser (Perfect Real Time) reverse transcription kit (Takara, Dalian, China).

For the quantitative PCR (qPCR) assay, the 25 µl reaction system was established (Table S1). The relative expression varieties of tested genes were calculated with the $2^{-\Delta\Delta Ct}$ method [5]. The Ct value was corrected using the actin gene as the internal reference gene [25].

The one-way ANOVA method was used to identify significant differences (at 0.05 and 0.01 probability levels) between the means of the tested samples via the Graphpad Prism program (v. 8.3.0) for statistical analysis.

RESULTS AND DISCUSSION

Identification of BBX proteins in *G. max* genomic sequences

Although the *BBX* gene family has been identified in several plants [4], its function and evolution in *G. max* still need to be determined. In this study, a total of 26 protein sequences containing the B-box domain were obtained and further validated by querying the NCBI CDD and SMART databases. The results showed that all 26 sequences had the B-box domain, indicating that they were all members of the soybean BBX family (Fig. 1). According to the chromosome names and chromosomal positions, these 26 proteins with the B-box domain were renamed as GmBBX1-GmBBX26, respectively (Table S2).

The basic physical and chemical characteristics of GmBBX proteins are listed in Table S2. As shown, the protein sequences of GmBBXs varied in length from 193–392 with an average amino acid number of 293.5, indicating a MW range from 21,374.49 to 44,287.7 Da. GmBBX10 (193 aa) was the smallest protein, while GmBBX22 (392 aa) was the largest. The *pI* value of 21 GmBBXs was below 7.0 with a range of 4.76 (GmBBX19) to 8.9 (GmBBX22).

The results of the predicted subcellular localization showed that most GmBBX proteins (24 in number) were predicted to localize to the nucleus. In contrast, GmBBX3 was predicted to localize in the membrane system for exiting TM (transmembrane) structure (Table S1). Another exception was GmBBX16, which was putatively located in the chloroplast due to its signal peptide sequence. The localization results for the GmBBX proteins indicate that they mainly fulfill their functions in the nucleus, similar to *A. thaliana* BBXs [2]. Previous research also identified plant BBX proteins with TM regions such as ZmBBX3 and ZmBBX30 [15], but the signal peptide has yet to be found in plant BBX proteins.

The hydrophilic analysis showed that all GmBBX proteins had negative hydrophilic scores with the most hydrophilic protein being GmBBX3 with a GRAVY score of −0.145. The most hydrophobic protein was GmBBX11 with a GRAVY score of −0.657. Analysis of the instability index showed that 26 proteins had scores above 40 with the highest being 69.65, indicating that most were short-lived proteins.

Conserved domains and motifs in GmBBX proteins

Querying the CDD and SMART databases revealed that 24 GmBBX proteins contained a dual B-box domain in tandem repeats, both of which were classified as class I B-box structural domains (Fig. 1). The two exceptions were GmBBX5 and GmBBX8, which only contained a class I B-box structural domain. Interestingly, 8 proteins have the CCT (CO, CO-like, and TOC1) domain simultaneously. The results showed that the sequences of GmBBX were highly conserved, particularly in functional motifs such as B-box and CCT motifs (Fig. 1), which are similar to other plant BBX proteins [5].

The CCT domain plays a crucial role in transcriptional regulation and nuclear protein transport [5]. Therefore, the presence of CCT in these 8 GmBBX proteins suggests their function in regulating specific gene transcription [2].

In addition, 19 GmBBXs contained 1–3 copies of the low complexity region (LCR). It has been evidenced that LCR contributes to the media protein-protein interaction (PPI) [26]. The abundance of LCR indicates that the GmBBX proteins participate in complex PPIs. Previous research showed that the VALINE-PROLINE (VP) motif also had functions on PPI [3], but this motif was absent in GmBBXs. Four different types of GmBBX proteins have been discovered overall based on the arrangement of the domains. These include BBXs with a B-box domain, two B-box domains, a B-box and an additional CCT domain, and two B-boxes and contain an additional CCT domain (Fig. 1) [3]. In soybean, the number of BBX members was 0, 2, 18, and 6, while in tomato, it was 8, 5, 10, and 6, respectively. In pear, the corresponding numbers were 6, 4, 9, and 7 [27]. These results indicated that the plant BBX genes were independently expanded from a common ancestor.

Next, we analyzed the motifs dispersed in the GmBBX protein sequences to uncover their functional information (Fig. 1). The results showed that motifs 2 and 1 were the most conserved motifs that duplicated in all GmBBX proteins. This motif represents the signature of the B-box1 domain with the pattern of C-X2-C-X7-8-C-X2-D-X-A-X-L-C-X2-C-D-X3-H (Fig. 2) [2]. However, there is an exception for GmBBX1 (KRH76689), which has an extra peptide sequence with 26 aa in length inserted into motif 1. HHpred searching showed that this sequence was homologous to the partial sequence of the domain DUF2770 (Domain of unknown function 2770). The motif 10 was found in 8 sequences in the CCT domain (R-X5-R-Y-X2-K-X3-R-X3-K-X2-R-Y-X2-R-KX2-A-X2-R-X2-R-X2-G-R-F-X-K) (Fig. 2) [26]. The remaining motifs were predicted to have weak homologous sequences deposited in the PFAM database. In this regard, motif 9 was shared by 19 GmBBX proteins with a low homologous sequence to an 8-amino acid sequence in the DUF3746 domain. This result indicated that the sequences of GmBBXs have distinct characteristics.

Fig. 1 Gene structure and conservative motif analysis of the GmBBX proteins. (a) The motif compositions of GmBBX proteins. (b) The domain composition and exon-intron structure of GmBBX proteins. The green box indicates the 5′ and 3′ non-translational region, while the yellow box represents the coding region, and the black line indicates the introns.

Chromosome location and gene structure of *GmBBXs*

Chromosomal distribution based on the genomic location of 26 *GmBBX* genes was visualized. The results showed that these 26 *GmBBX* genes were unevenly distributed on 12 soybean chromosomes, of which chromosomes 11 and 13 had the highest number of *GmBBX* genes. These 2 chromosomes were the hotspots of the *GmBBX* gene distribution. In addition, chromosomes 1, 3, 9, and 15 had the lowest number of *GmBBX* genes, all with only one.

To assess the evolutionary relationships among members of the *GmBBX* family, we analyzed their structure diversity. The exon/intron organization of sequences reflects the structural diversity and complexity of genes within the same gene family [28]. The results showed that the 26 *GmBBX* genes contained 2–5 exons and 1–5 introns. Of them, the *GmBBX3*, *GmBBX5*, *GmBBX7*, *GmBBX8*, *GmBBX17*, *GmBBX18*, *GmBBX23*, and *GmBBX25* only have 2 exons, while *GmBBX1*, *GmBBX10*, *GmBBX11*, and *GmBBX14* have 5 exons. Specifically, it showed that the closer the genetic relationship, the more similar the genetic structure between them.

Multiple sequence alignment and phylogenetic evolution of GmBBXs

In addition to the gene structure analysis, the unroot phylogenetic tree analysis was performed on 111 BBX proteins from 7 different plant species to investigate

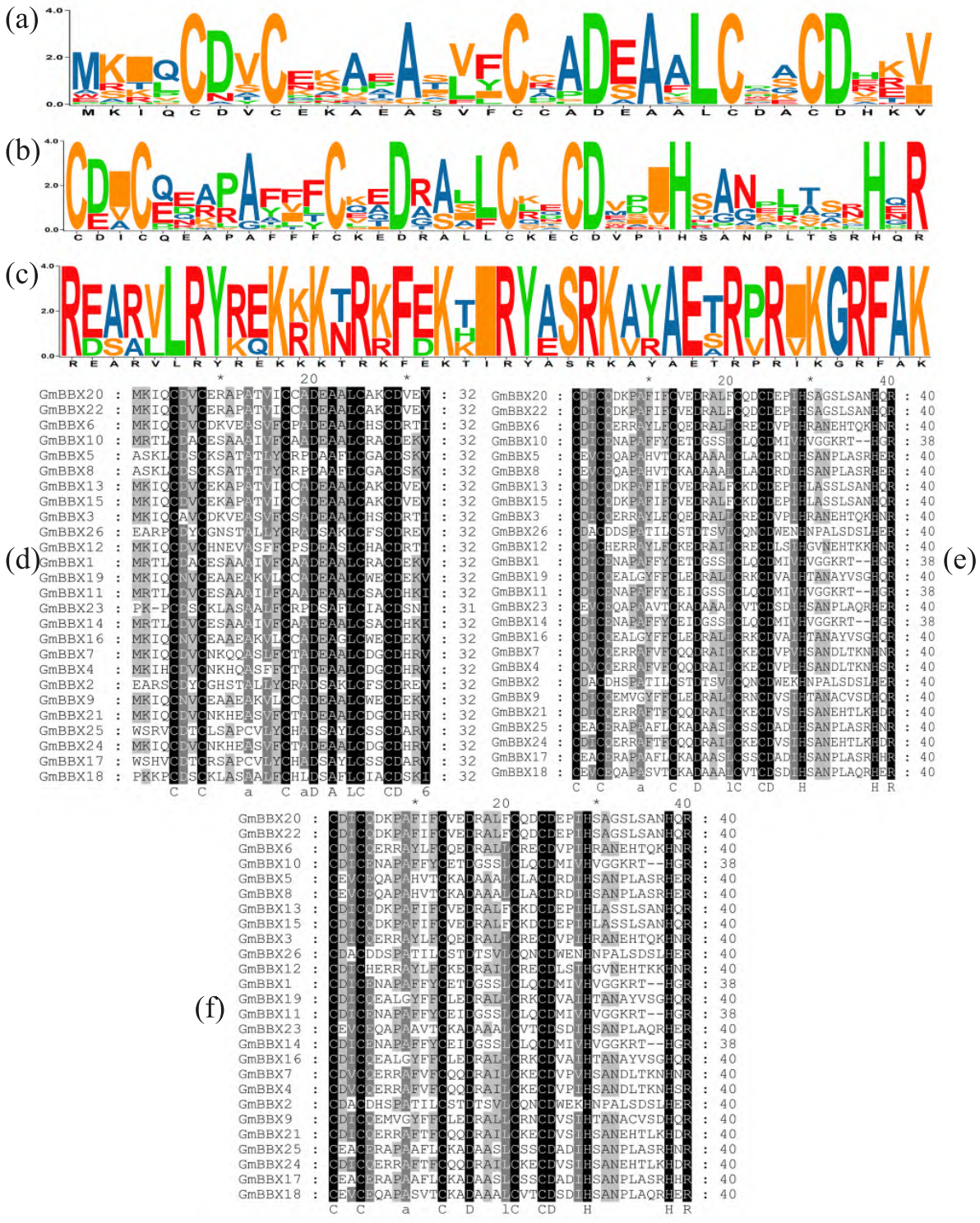

Fig. 2 Seqlogo of 3 motifs and sequence alignments of B-box domains and CCT domains. (a–c) Seqlogo of motif 1, motif 2 and motif 10, respectively; (d–f) sequences alignment of B-box 1 domain, B-box 2 domain, and CCT domain, respectively.

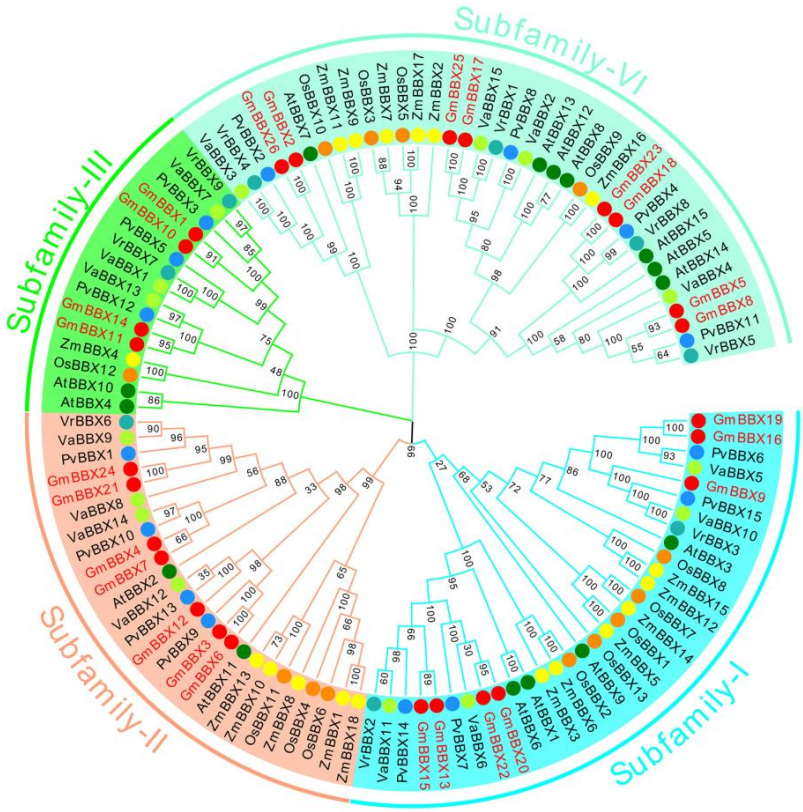

Fig. 3 Phylogenetic tree of the BBX proteins from 7 plant species. Gm: *G. max*, Vr: *V. radiata*, Pv: *P. vulgaris*, Zm: *Z. mays*, Os: *O. sativa*, At: *A. thaliana*, and Va: *V. angularis*.

the evolutionary relationships between these BBX proteins across these different plant species. The results showed that these BBX proteins could be divided into 4 families, classified as subfamily I to IV. Among them, subfamily IV was the largest group with 37 BBX proteins, while subfamily III was the smallest with 16 members (Fig. 3). These results revealed that groups III and IV expanded more than the other 2 groups.

The distribution of GmBBX proteins across the 4 subfamilies was found to be uneven. Specifically, subfamily IV consisted of 8 GmBBXs that possessed the CCT domain, and other members in this subfamily such as VaBBX3, VrBBX5, and AtBBX5 also contained CCT domains (Fig. 3). Within each subfamily, GmBBXs demonstrated a close clustering pattern, subsequently grouping with proteins from *P. vulgaris*, *V. radiata*, and *V. angularis*. These findings suggest a close evolutionary relationship between *G. max* and these plant species. In contrast, the BBXs from *O. sativa* and *Z. mays* were closely grouped, indicating a distant evolutionary relationship between these monocot plants and dicot plants.

Previously, it was believed that B-box1 domain was more conserved than B-box2 domain, and the deletion events were more frequently observed in the B-box2 domain. We also found that B-box1 was more conserved than the B-box2 domain, indicating that the deletion process could occur in B-box2 domains during evolution (Fig. 2).

Additionally, some of the clustered *GmBBXs* shared similar gene architectures. For instance, the gene architectures of *GmBBX1* and *GmBBX10* are identical, both including the same number of exons. In certain instances, it was found that gene pairs with similar gene structure are not positioned in close proximity to each other on the chromosome. This is exemplified by the case of *GmBBX11* and *GmBBX14*, which possess similar numbers of exons and introns but are located on separate chromosomes, namely chromosomes 11 and 12, respectively.

Synteny and gene duplication analysis

Gene duplications, mainly in the form of tandem and segmental duplication, are believed to drive gene family expansion [29]. An initial collinearity analysis was conducted to investigate gene duplication occurrences in the *GmBBX* genes. The results revealed a total of 15 duplicated gene pairs linked to 18 *GmBBXs* with

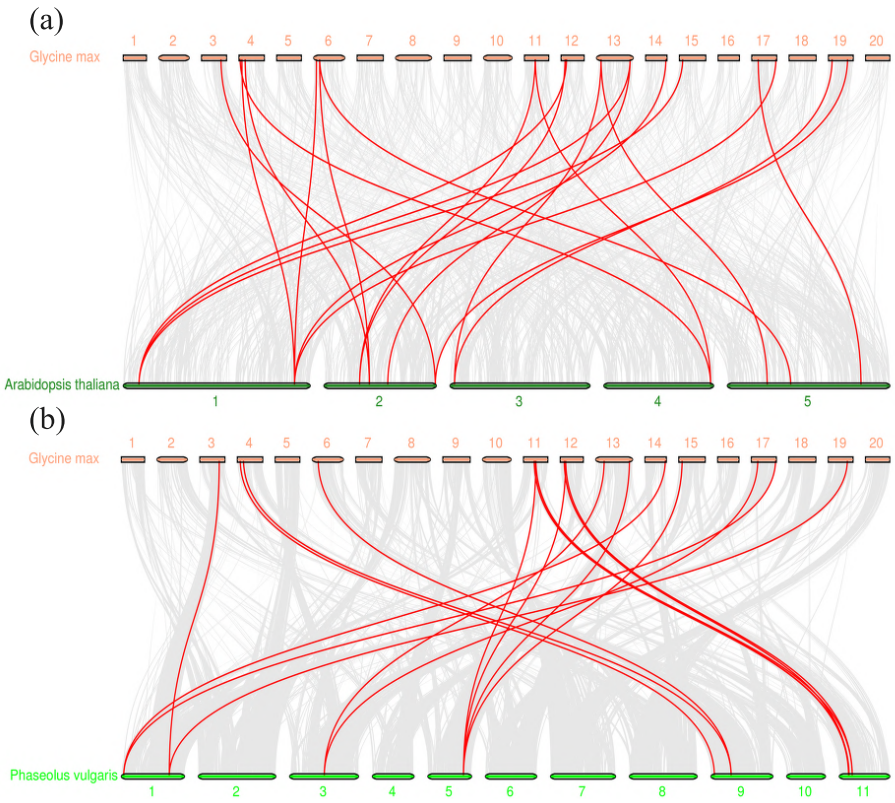

Fig. 4 Collinearity relationships of GmBBX proteins to AtBBX and PvBBX proteins. The grey lines indicated collinearity relationships among the soybean proteins, while the red lines referred to the duplicated GmBBX protein pairs. (a) Results from collinearity relationships analysis between *G. max* and *A. thaliana* BBX proteins. (b) The corresponding results of *G. max* and *P. vulgaris* BBX proteins. The numbers under the yellow and green lines represent the *G. max* and the plant chromosome numbers used for collinearity analysis of proteins with *G. max*.

each pair being classified as a segmental duplication. Notably, the majority of segmental duplication events were attributed to *GmBBXs* on chromosomes 4, 13, and 17. Chromosomes 6, 11, and 14 also contributed to the duplication events. The proteins contributing to duplication events in the *GmBBX* gene family are *GmBBX21* (KRH17384) and *GmBBX24* (KRH05902). In conclusion, most *GmBBXs* may have originated from gene duplications, and segmental duplication events may play a crucial role in generating new *GmBBX* genes. The duplicated genes without tandem duplication events were also revealed in the BBX family genes from maize, rice, and sorghum [15].

It has been well documented that structural divergence in duplicated genes has been widespread and, in many cases, led to the generation of functionally distinct paralogs [30]. The results in this study clearly showed a predominant similarity in gene structure among most paralogous pairs, particularly in terms of exon number and sequence length. For example, among the 15 paralogous gene pairs, 11 showed highly consistent exon organization, including exon numbers and lengths. However, there were notable differences in the lengths of introns and untranslated regions (UTR) related to gene expression regulation [31]. Furthermore, gene pairs without similar gene structures were discovered in this research. In the gene pairs *GmBBX22/GmBBX20* (KRH10122/KRH232140), *GmBBX21/GmBBX7* (KRH17384/KRH51757), and *GmBBX24/GmBBX7* (KRH05902/ KRH51757), the gene structure was greatly varied. These results showed that the evolutionary history of these paralogs was diverse. Previous studies have reported that segmental duplication usually occurs in a gene family that is slowly evolving [31]. The present results thus indicate that *GmBBX* is a gene family with slow evolutionary characteristics.

We conducted 3 comparative syntenic analyses between *G. max* and 3 representative species, including 2 dicots, *P. vulgaris* and *A. thaliana*, and one monocot, *Z. mays*, to uncover the evolutionary clues for the GmBBXs (Fig. 4). The most common synteny was found between AtBBXs and GmBBXs, followed by PvBBXs. Nineteen GmBBXs and 7 AtBBXs formed a total of 42 syntenic relationships, as depicted in

Fig. 4a. Similarly, 17 GmBBXs and 9 PvBBXs formed 36 syntenic relationships. Unlike AtBBXs and PvBBXs, BBXs originated from monocot *Z. mays* (ZmBBX) did not exhibit synteny with GmBBXs. A total of 16 GmBBXs were involved in syntenic gene pairs with those of AtBBXs and PvBBXs at the same time (Fig. 4b). Among these GmBBX proteins, GmBBX24 (KRH05902) and GmBBX20 (KRH23214) exhibited the highest number of relationships with other BBXs, each having 5 connections. This was followed by GmBBX21 (KRH17384), GmBBX15 (KRH24617), and GmBBX8 (KRH52295), each demonstrating 4 relationships. These results provide valuable insights into the evolution of the plant BBX family.

To better understand the history of coding sequence selection affecting the *GmBBX* gene coding sequences, we calculated the ratios of nonsynonymous (Ka) versus synonymous (Ks) mutation of orthologous gene pairs. As shown in Table S3, the Ka/Ks ratios of 15 *GmBBX* orthologous gene pairs were significantly lower than 1, suggesting that these pairs faced purifying selection pressure during evolution [32]. Based on the divergence rate of 6.1×10^{-9} synonymous mutations per synonymous site year proposed for soybean [30], we estimated the timing of duplicating events of paralogous *GmBBX* gene pairs. The results showed that segmental duplication events occurred between 7.59 and 86.60 Mya.

Cis-element analyses of *GmBBX* genes

As binding sites for TFs, the *cis*-elements within promoter regions play a crucial role in gene expression regulation [33]. To analyze the expression regulation patterns of the *GmBBX* genes, we extracted the 2,000-bp upstream sequences of 26 *GmBBX* genes and identified the *cis*-acting elements. Our findings revealed that the most prevalent groups among the 1,151 associated *cis*-acting elements were related to light responsiveness, phytohormone responsiveness, stress responsiveness, and plant growth and development (Fig. 5). As a result, it was believed that various factors could influence the expression of *GmBBX* genes. More specifically, 12 different *cis*-elements were found to be responsible for phytohormones, including abscisic acid (ABA), salicylic acid (SA), auxin, methyl jasmonate (MeJA), ethylene (ET), and gibberellin. ABA-responsive element (ABRE), ethylene-responsive element (ERE), CGTCA-motif, TGACG-motif, TCA, AuxRR-core, P-box, and TGA-element were among the *cis*-elements (Fig. 5). Phytohormones are a class of small signaling molecules with multiple functions. Among them, ABA, SA, MeJA, and ET play important roles in responding to various abiotic stresses. Therefore, the presence of *cis*-elements relevant to transcriptional regulation by various phytohormones suggests that *GmBBX* members may fulfill anti-abiotic and anti-biotic stress functions. In addition, a total of 13 types of *cis*-elements were assigned to the abiotic stress responsiveness group. In this group, the MYC binding element and anaerobic responsiveness element (ARE) were the most abundant elements, followed by MYB-binding site (MBS) [34, 35].

The presentation of these elements, also including ABRE and ERE, indicated the important roles of *GmBBX* during adaptation to abiotic stress. The expression of plant *BBX* genes have been evidenced responding to abiotic stress and hormonal treatments like ABA [36, 37]. Recently, gene co-expression network analysis conducted under UV-B radiation revealed a significant association between the expression of *Vaccinium corymbosum* BBXs (*VcBBXs*) and 41 differentially expressed genes (DEGs) related to the ET signaling pathway as well as 7 DEGs associated with the ABA signaling pathway [36]. In another study, Li et al[37] examined the expression changes of *Salvia miltiorrhiza* BBXs (*SmBBXs*) under ABA, PEG, and NaCl treatments. Under ABA treatment, the expression levels of *SmBBX 5, 6, 12, 20, 21, 22, 23*, and *27* were elevated, while the expression of other genes decreased. Similar observations were made with NaCl treatment; *SmBBX 5, 26, 20, 25*, and *22* showed increased expression.

Notably, the *GmBBX* genes, which were identified as paralogous genes, shared a similar constitution in their *cis*-element promoter regions. For *GmBBX4/GmBBX7* gene pairs, there were 15 *cis*-elements present in the promoters. In addition, this phenomenon was also observed in *GmBBX7/GmBBX21* gene pairs, highlighting the conserved evolutionary pattern of both the coding and promoter sequences.

Expression profiling of *GmBBX* genes

In order to characterize the expression pattern of *GmBBX* genes across different tissues and developmental periods, we extract transcriptome data from 5 tissues, including flower, seed, shoot, nodules, and leaf.

Most *GmBBX* genes were expressed at low levels. The highest expression values for *GmBBX 20* and *22* were 239.416 and 201.746, respectively. *GmBBX 20* and *22* genes were grouped in the same subcluster I, suggesting that some genes with a closed evolutionary relationship had a similar expression profile. This phenomenon was also found for gene pairs *GmBBX13/GmBBX15* and *GmBBX5/GmBBX8* but not for *GmBBX19 and GmBBX16*.

We further analyzed the *cis*-elements within the promoters of *GmBBX20* and *GmBBX22* and found that the elements responding to light and abiotic stress were abundant. In the former group, Box 4, G-box, I box, and TCT motifs appeared 18 and 20 times in their promoters, respectively, while the ABRE, ERE, dehydration-responsive element (DRE), and wound-responsive element 3 (WRE3) were present several

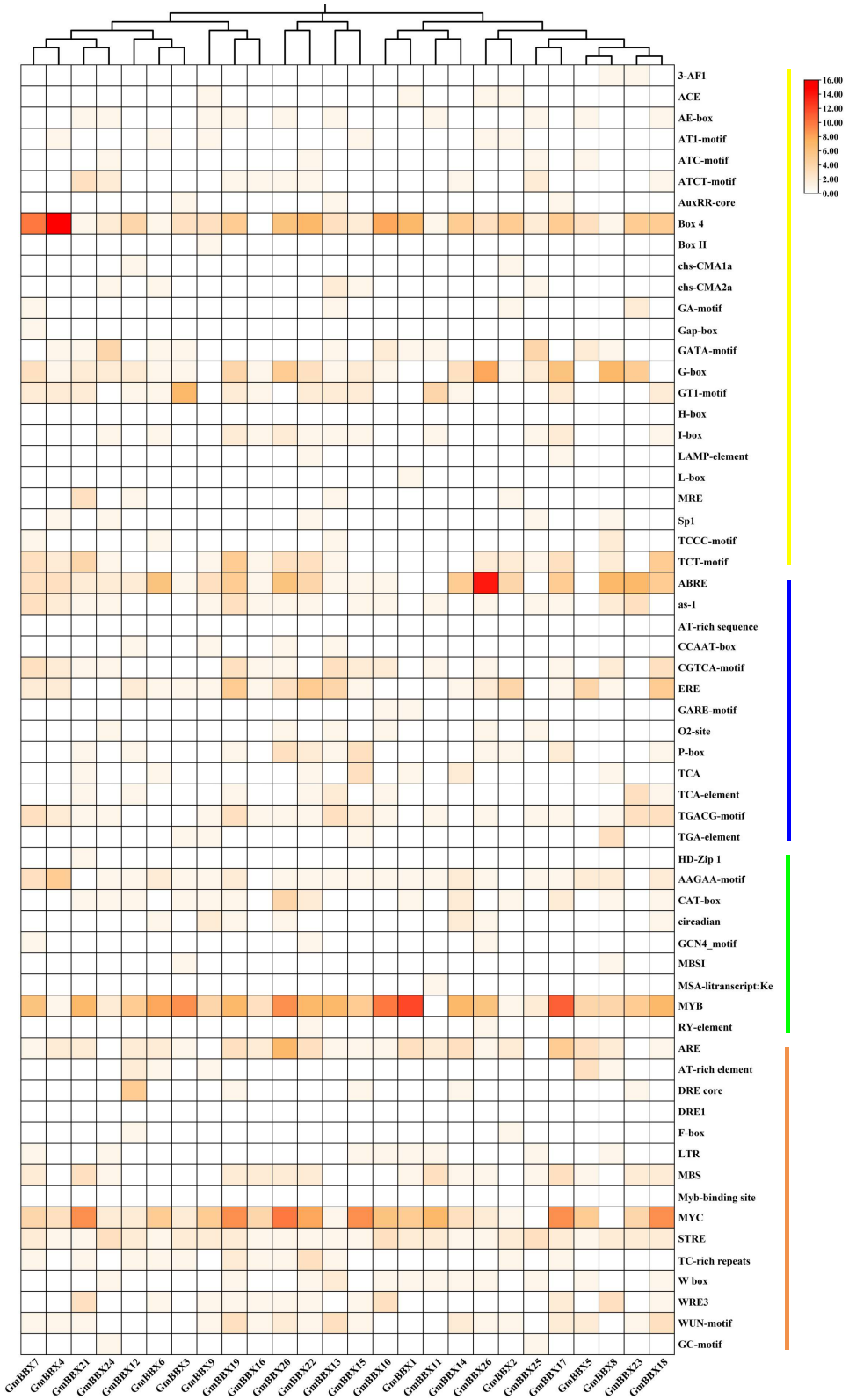

Fig. 5 The *cis*-element compositions in promoter sequences of *GmBBX* genes. The red line indicated the elements assigned to abiotic responsiveness, while the green lines to development and growth responsiveness, the blue line to phytohormone responsiveness, and the yellow line to light responsiveness.

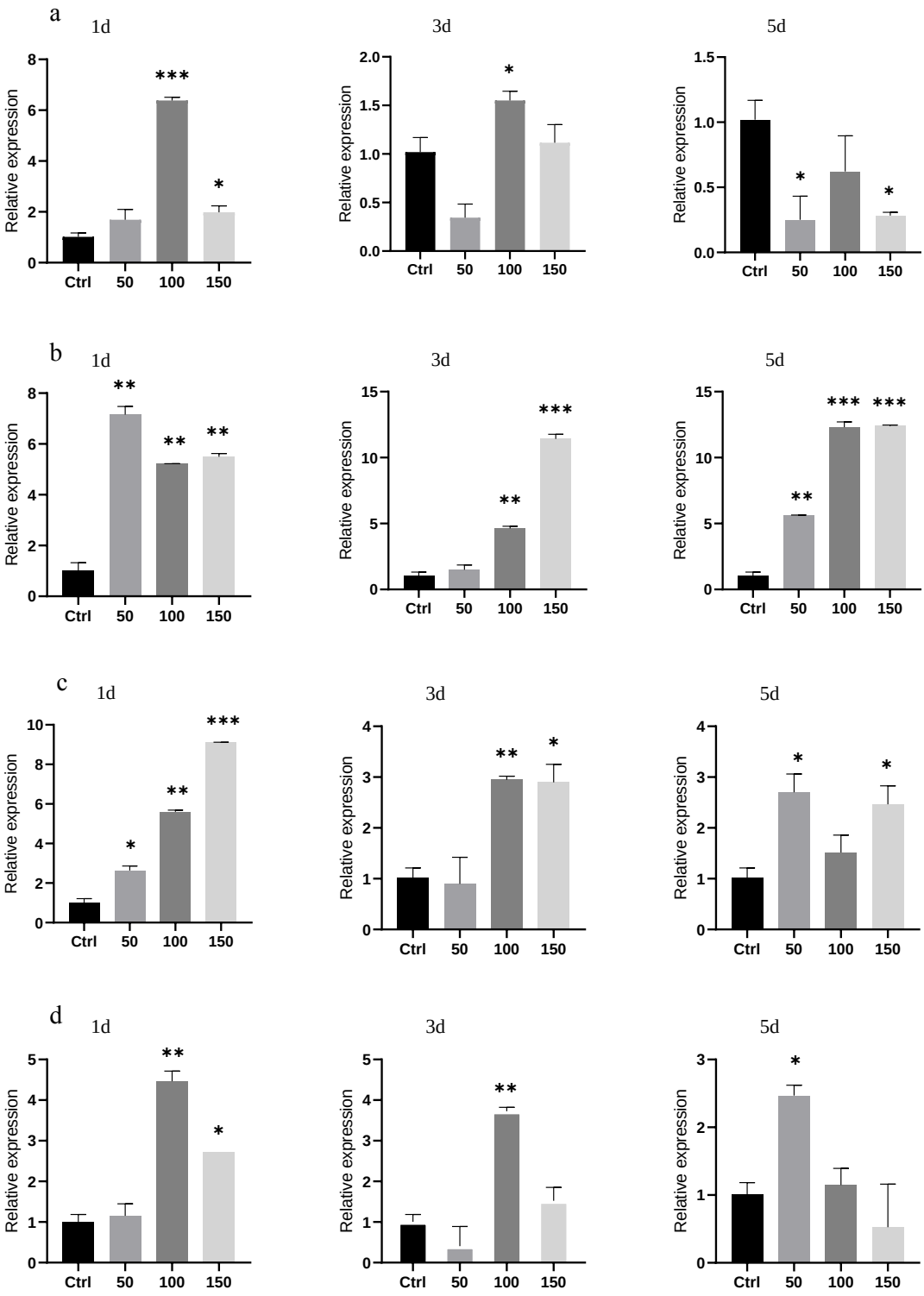

Fig. 6 Expression of 4 *GmBBX* genes under different concentrations of NaCl (0, 50, 100, and 150 mM). (a) *GmBBX18* (KRH18980), (b) *GmBBX20* (KRH23214), (c) *GmBBX22* (KRH10122), and (d) *GmBBX23* (KRH02915). One to three stars represent significant differences in the analysis with p-values less than 0.05, 0.01, and 0.001, respectively. Various treatment groups were compared with the control group.

times (Fig. 5). These results suggest that these *GmBBX* genes may be important genes under abiotic stress.

Based on expression levels and *cis*-element constitution, 4 *GmBBX* genes, including *GmBBX 18, 20, 22*, and *23*, were selected for qPCR analysis under 0, 50, 100, and 150 mM NaCl treatment for 1 to 5 days (Fig. 6). The results showed a significant fluctuation in expression levels, which depended on the stress conditions and stress periods. In this regard, the expression profile of these 4 genes significantly increased under NaCl stress for 1 day ($p < 0.001$ or $p < 0.01$), while it changed significantly after 3 and 5 days of NaCl stress (Fig. 6). For *GmBBX18* (KRH18980), its expression showed a significant increase at 1 and 3 days of NaCl stress, compared to control, with a significant increase at 50 and 100 mM NaCl stress. However, a dramatic decrease in expression was observed at all 3 concentrations after 5 days of salt stress. For *GmBBX20* (KRH23214), its expressional trend was similar to that of *GmBBX18* with the exception that its expression increased 6.25-fold at 150 mM salt stress for 5 days. For *GmBBX 23* and *22*, their expression levels also showed different degrees of increase after 1 day of salt stress, and their highest expression levels occurred at 100 mM salt stress for 3 days with an increase of 1.78- and 3.01-fold, respectively.

At the same time, we also analyzed the change of 4 morphological parameters including fresh weight, dry weight, plant height, and leaf area. The results showed that both PEG and NaCl treatments decreased these parameters in a dose and time dependent manner (Fig. S1), implying that these 2 abiotic stresses have an adverse impact on plant growth. In contrast, the expression of the 2 selected *GmBBX* genes, *GmBBX20* (KRH23214) and *GmBBX22* (KRH10122), increased significantly, while other expressional patterns of the 2 genes *GmBBX18* (KRH18980) and *GmBBX23* (KRH02915) decreased followed by an increase after treatment for 5 days. These results suggested that several *GmBBX* genes might have important functions under NaCl stress, as evidenced by previous research [3, 12, 15].

CONCLUSION

In this study, a genome-wide identification and characterization of the *BBX* gene family including their evolutionary relationships, protein and gene structures, and expression patterns, was performed in soybean. The 26 GmBBX proteins have been classified into 4 subfamilies. A total of 18 *GmBBXs* formed 15 segmental duplications in *G. max*, while 19 *GmBBX* and 11 *AtBBX* genes contributed to 42 cross-species syntenic relationships. The transcriptome data showed that the *GmBBXs* were mainly expressed in leaves and roots, and their expression profiles were altered under drought and NaCl stress. qPCR results of 4 chosen *GmBBX* genes showed a more pronounced rise in gene expression under 50–150 mM NaCl stress. In summary, the results indicated that the GmBBXs were conserved plant BBX proteins and putative key regulators involved in stress tolerance. These results provide valuable information for further functional characterization of *GmBBX* genes in stress response to improve soybean breeding.

Appendix A. Supplementary data

Supplementary data associated with this article can be found at https://dx.doi.org/10.2306/scienceasia1513-1874.2024.109.

REFERENCES

1. Chu Z, Wang X, Li Y, Yu H, Li J, Lu Y, Li H, Ouyang B (2016) Genomic organization, phylogenetic and expression analysis of the B-BOX gene family in tomato. *Front Plant Sci* **7**, 1552.
2. Gangappa SN, Botto JF (2014) The BBX family of plant transcription factors. *Trends Plant Sci* **19**, 460–470.
3. Talar U, Kiełbowicz-Matuk A (2021) Beyond *Arabidopsis*: BBX regulators in crop plants. *Int J Mol Sci* **22**, 2906.
4. Hou W, Ren L, Zhang Y, Sun H, Shi T, Gu Y, Wang A, Ma D, et al (2021) Characterization of BBX family genes and their expression profiles under various stresses in the sweet potato wild ancestor *Ipomoea trifida*. *Sci Hortic* **88**, 110374.
5. Ma R, Chen J, Huang B, Huang Z, Zhang Z (2021) The BBX gene family in Moso bamboo (*Phyllostachys edulis*): identification, characterization and expression profiles. *BMC Genomics* **22**, 533.
6. Zhao J, Li H, Huang J, Shi T, Meng Z, Chen Q, Deng J (2021) Genome-wide analysis of BBX gene family in Tartary buckwheat (*Fagopyrum tataricum*). *Peer J* **11**, e11939.
7. Min JH, Chung JS, Lee KH, Kim CS (2015) The CONSTANS-like 4 transcription factor, AtCOL4, positively regulates abiotic stress tolerance through an abscisic acid-dependent manner in Arabidopsis. *J Integr Plant Biol* **57**, 313–324.
8. Yang Y, Ma C, Xu Y, Wei Q, Imtiaz M, Lan H, Gao S, Cheng L, et al (2014) A zinc finger protein regulates flowering time and abiotic stress tolerance in *Chrysanthemum* by modulating gibberellin biosynthesis. *Plant Cell* **26**, 2038–2054.
9. Zhu C, Xiao L, Hu Y, Liu L, Liu H, Hu Z, Liu S, Zhou Y (2022) Genome-wide survey and expression analysis of B-Box family genes in cucumber reveal their potential roles in response to diverse abiotic and biotic stresses. *Agriculture* **12**, 827.
10. Wang Q, Tu X, Zhang J, Chen L, Rao L (2013) Heat stress-induced BBX18 negatively regulates the thermotolerance in *Arabidopsis*. *Mol Biol Rep* **40**, 2679–2688.
11. Wang Y, Zhang Y, Liu Q, Zhang T, Chong X, Yuan H (2021) Genome-wide identification and expression analysis of BBX transcription factors in *Iris germanica* L. *Int J Mol Sci* **22**, 8793.
12. Wu H, Wang X, Cao Y, Zhang H, Hua R, Liu H, Sui S (2021) CpBBX19, a B-Box transcription factor gene of *Chimonanthus praecox*, improves salt and drought tolerance in *Arabidopsis*. *Genes* **12**, 1456.

13. Poonpakdee C, Ntlopo KZ, Onthong J, Khawmee K, Lin YT (2023) Response and efficiency of magnesium fertilizer application in soybean (*Glycine max*) and sunflower (*Helianthus annuus*). *ScienceAsia* **49**, 462–468.
14. Hossain Z, Khatoon A, Komatsu S (2013) Soybean proteomics for unraveling abiotic stress response mechanism. *J Prot Res* **12**, 4670–4684.
15. Shalmani A, Jing XQ, Shi Y, Muhammad I, Zhou MR, Wei XY, Chen QQ, Li WQ, et al (2019) Characterization of B-BOX gene family and their expression profiles under hormonal, abiotic and metal stresses in *Poaceae* plants. *BMC Genomics* **20**, 27.
16. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* **13**, 1194–1202.
17. Su W, Ren Y, Wang D, Huang L, Fu X, Ling H, Su Y, Huang N, et al (2020) New insights into the evolution and functional divergence of the CIPK gene family in *Saccharum*. *BMC Genomics* **21**, 868.
18. Li Y, Liu Z, Zhang K, Chen S, Liu M, Zhang Q (2020) Genome-wide analysis and comparison of the DNA-binding one zinc finger gene family in diploid and tetraploid cotton (*Gossypium*). *PLoS One* **15**, e0235317.
19. Patil G, Valliyodan B, Deshmukh R, Prince S, Nicander B, Zhao M, Sonah H, Song L, et al (2015) Soybean (*Glycine max*) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics* **16**, 520.
20. Quan S, Niu J, Zhou L, Xu H, Ma L, Qin Y (2019) Genome-wide identification, classification, expression and duplication analysis of GRAS family genes in *Juglans regia* L. *Sci Rep* **9**, 11643.
21. Belamkar V, Weeks NT, Bharti AK, Farmer AD, Graham MA, Cannon SB (2014) Comprehensive characterization and RNA-Seq profiling of the HD-Zip transcription factor family in soybean (*Glycine max*) during dehydration and salt stress. *BMC Genomics* **15**, 950.
22. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527.
23. Li Y, Andrade J (2017) DEApp: an interactive web interface for differential expression analysis of next generation sequence data. *Source Code Biol Med* **12**, 2.
24. Nahar K, Hasanuzzaman M, Suzuki T, Fujita M (2017) Polyamines-induced aluminum tolerance in mung bean: A study on antioxidant defense and methylglyoxal detoxification systems. *Ecotoxicology* **26**, 58–73.
25. Nakayama TJ, Rodrigues FA, Neumaier N, Marcelino-Guimarães FC, Farias JR, de Oliveira MC, Borém A, de Oliveira AC, et al (2014) Reference genes for quantitative real-time polymerase chain reaction studies in soybean plants under hypoxic conditions. *Genet Mol Res* **13**, 860–871.
26. Obel HO, Cheng C, Li Y, Tian Z, Njogu MK, Li J, Lou Q, Yu X, et al (2022) Genome-wide identification of the B-Box gene family and expression analysis suggests their potential role in photoperiod-mediated β-carotene accumulation in the endocarp of cucumber (*Cucumis sativus* L.) fruit. *Genes* **13**, 658.
27. Cao Y, Han Y, Meng D, Li D, Jiao C, Jin Q, Lin Y, Cai Y (2017) B-BOX genes: genome-wide identification, evolution and their contribution to pollen growth in pear (*Pyrus bretschneideri* Rehd). *BMC Plant Biol* **17**, 156.
28. Xu G, Guo C, Shan H, Kong H (2012) Divergence of duplicate genes in exon-intron structure. *Proc Natl Acad Sci USA* **109**, 1187–1192.
29. Wang L, Ding X, Gao Y, Yang S (2020) Genome-wide identification and characterization of GRAS genes in soybean (*Glycine max*). *BMC Plant Biol* **20**, 415.
30. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155.
31. Chung BY, Simons C, Firth AE, Brown CM, Hellens RP (2006) Effect of 5'UTR introns on gene expression in *Arabidopsis thaliana*. *BMC Genomics* **7**, 120.
32. Wang X, Wu MH, Xiao D, Huang RL, Zhan J, Wang AQ, He LF (2021) Genome-wide identification and evolutionary analysis of RLKs involved in the response to aluminum stress in peanut. *BMC Plant Biol* **21**, 281.
33. Inukai S, Kock KH, Bulyk ML (2017) Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev* **43**, 110–119.
34. Bhuria M, Goel P, Kumar S, Singh AK (2016) The promoter of AtUSP is co-regulated by phytohormones and abiotic stresses in *Arabidopsis thaliana*. *Front Plant Sci* **7**, 1957.
35. Cui X, Zhang P, Hu Y, Chen C, Liu Q, Guan P, Zhang J (2021) Genome-wide analysis of the universal stress protein A gene family in *Vitis* and expression in response to abiotic stress. *Plant Physiol Biochem* **165**, 57–70.
36. Liu X, Sun W, Ma B, Song Y, Guo Q, Zhou L, Wu K, Zhang X, et al (2023) Genome-wide analysis of blueberry B-box family genes and identification of members activated by abiotic stress. *BMC Genomics* **24**, 584.
37. Li Y, Tong Y, Ye J, Zhang C, Li B, Hu S, Xue X, Tian Q, et al (2023). Genome-wide characterization of B-Box gene family in *Salvia miltiorrhiza*. *Int J Mol Sci* **24**, 2146.

Appendix A. Supplementary data

Table S1 The detailed information of qPCR analysis.

Reagent	Volume added	qPCR parameter
TB Green Premix Ex Taq II	12.5 µl	95 °C 2 min
primers	Each 1.0 µl	94 °C, 30 s
cDNA template	1 µl	57 °C, 20 s
nuclease-free water	9.5 µl	72 °C, 20 s
Final volume	25 µl	40 cycles

Genes	Primer sequences F	Primer sequences R	PL[a]
GmBBX20 (KRH23214)	ATCTTTGCATCCCCACCAGCTC	ACGAACCCTTTACCAGACACGGTA	224
GmBBX22 (KRH10122)	TTGTTGCGCGGATGAGGCTG	GCAAGGAAGCGCTGGTGGTT	249
GmBBX18 (KRH18980)	TTGTGGTGTCGGGTGGACAA	GACGCTCCACATCCGAGTCA	204
GmBBX23 (KRH02915)	TAGTGTGGGGCAAGGTGCGA	GTCCAGCATGAGCACCGCAG	232
actin	TGGCCATCCAAGCTGTCCTCT	GGCATGAGGGAGTGCATAGCC	122

[a] PL, Products length (base pair, bp).

Table S2 Physical and chemical properties of GmBBX proteins.

B-box ID	Protein ID	Length	MW	*pI*	Insta[a]	GRAVY	Chr	+/−	Localization	SL[b]
GmBBX1	KRH76689	210	23223.74	6.69	51.98	–0.378	1	+	50609992–50613102	nuclear
GmBBX2	KRH68113	349	38811.41	5.98	57.19	–0.617	3	–	41649707–41654690	nuclear
GmBBX3	KRH60788	319	36520.66	7.83	44.90	–0.145	4	–	734314–736030	membrane
GmBBX4	KRH61078	266	29383.00	6.58	69.65	–0.569	4	–	2193861–2195469	nuclear
GmBBX5	KRH61628	309	33732.84	7.01	50.73	–0.409	4	+	4830134–4831900	nuclear
GmBBX6	KRH51476	233	26123.20	6.49	45.49	–0.574	6	–	715829–717010	nuclear
GmBBX7	KRH51757	245	27171.57	6.11	55.50	–0.440	6	–	2087092–2088456	nuclear
GmBBX8	KRH52295	310	33958.11	7.53	52.81	–0.428	6	+	4519732–4521199	nuclear
GmBBX9	KRH37922	292	31989.71	5.37	60.92	–0.464	9	–	17582481–17586119	nuclear
GmBBX10	KRH28766	193	21374.49	7.01	50.49	–0.544	11	–	5593535–5596130	nuclear
GmBBX11	KRH29344	212	23765.85	5.90	63.95	–0.657	11	–	8475007–8478420	nuclear
GmBBX12	KRH29378	288	31836.76	6.34	58.52	–0.454	11	–	8627210–8628706	nuclear
GmBBX13	KRH29616	238	26080.60	5.01	42.64	–0.282	11	+	9656242–9659517	nuclear
GmBBX14	KRH24373	212	23682.94	6.70	60.11	–0.592	12	+	2698742–2702408	nuclear
GmBBX15	KRH24617	238	26083.70	5.10	37.70	–0.273	12	+	3704987–3708481	nuclear
GmBBX16	KRH27404	374	41033.16	4.97	51.83	–0.207	12	–	39321100–39325411	membrane/chloroplast
GmBBX17	KRH18308	361	40218.75	5.32	44.78	–0.492	13	+	14630394–14632571	nuclear
GmBBX18	KRH18980	365	39873.84	6.21	50.74	–0.332	13	–	20903249–20905141	nuclear
GmBBX19	KRH21881	293	31955.77	4.76	56.94	–0.330	13	+	36804691–36808953	nuclear
GmBBX20	KRH23214	239	26362.72	5.00	43.56	–0.321	13	–	43563135–43565875	nuclear
GmBBX21	KRH17384	276	30556.32	6.30	51.13	–0.413	14	–	48144382–48146183	nuclear
GmBBX22	KRH10122	392	44287.70	8.90	55.65	–0.461	15	+	2377565–2381868	nuclear
GmBBX23	KRH02915	374	40706.58	5.94	48.05	–0.336	17	+	5146337–5148259	nuclear
GmBBX24	KRH05902	327	36454.13	6.15	53.67	–0.362	17	+	40923558–40925409	nuclear
GmBBX25	KRG93763	366	40905.64	5.47	47.50	–0.507	19	–	5467686–5469895	nuclear
GmBBX26	KRG96381	351	38947.59	5.88	55.68	–0.591	19	–	46241152–46245817	nuclear

[a] Instability index, [b] Subcellular localization.

Table S3 Ka/Ks ratios of the paralog *GmBBX* gene pairs.

Gene 1	Gene 2	Ka	Ks	Ka/Ks	Duplication rate (Mya)
KRH29378	KRH60788	0.370	1.057	0.350	86.600
KRH51757	KRH61078	0.058	0.141	0.410	11.545
KRH17384	KRH61078	0.289	0.695	0.417	56.950
KRH05902	KRH61078	0.282	0.771	0.366	63.168
KRH52295	KRH61628	0.015	0.308	0.048	25.204
KRH17384	KRH51757	0.277	0.699	0.397	57.308
KRH05902	KRH51757	0.259	0.725	0.358	59.381
KRH24373	KRH29344	0.021	0.093	0.222	7.590
KRH24617	KRH29616	0.028	0.167	0.167	13.651
KRH23214	KRH29616	0.089	0.510	0.173	41.832
KRH23214	KRH24617	0.079	0.466	0.170	38.169
KRG93763	KRH18308	0.029	0.153	0.187	12.546
KRH02915	KRH18980	0.030	0.293	0.102	24.045
KRH10122	KRH23214	0.013	0.102	0.126	8.334
KRH05902	KRH17384	0.035	0.151	0.230	12.3546

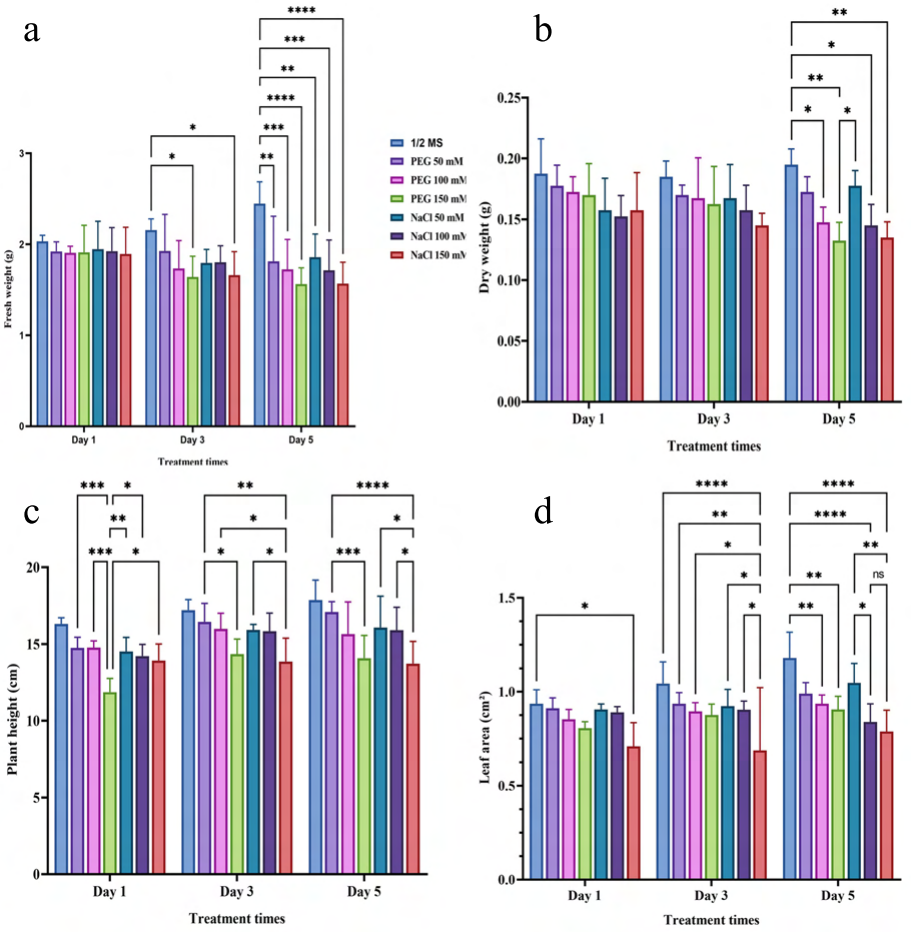

Fig. S1 The variations of 4 morphological parameters under PEG and NaCl treatments. One to four stars represent significant differences in the analysis with p-values less than 0.05, 0.01, 0.001, and 0.001, respectively.