

Genome-wide identification and expression analysis of the WRKY gene family in five legumes

Lei Ling (✉ linglei521lei@163.com)

Daqing Normal University

He Zhao

Daqing Normal University

Naiyu Chen

Daqing Normal University

Yajie Yin

Daqing Normal University

Hua Yue

Daqing Normal University

Shumian Qin

Daqing Normal University

Xiaoju Zhao

Daqing Normal University

Guoling Ren

Daqing Normal University

Research Article

Keywords: Legumes, whole genome, WRKY gene family, Abiotic stress

Posted Date: August 29th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1921086/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: *WRKY* is a transcription factor with 60 highly conserved amino acid region, regulating physiological processes and development in many plants and respond to a variety of abiotic stresses. However, the *WRKY* gene family has not been systematically studied in multiple legumes.

Results: In this study, genome-wide analysis of five legumes (*Phaseolus vulgaris*, *Medicago truncatula*, *Lotus japonicus*, *Glycine max*, *Cicer arietinum*) identified 90, 108, 83, 185 and 61 *WRKY* genes, respectively. Besides, the chromosomal localization, phylogenesis, conserved motifs, gene structure and gene replication analysis were proceed for *WRKY* genes. The results showed that the legumes *WRKY* gene branches from the same evolutionary system have similar protein structures. In addition, by analyzing the transcriptome data we figured out that the expression level of *WRKY* differed in each tissue. Meanwhile, qRT-PCR was carried out to investigate the expression patterns of *WRKY* under abiotic stress. The results show that *WRKY* protein play crucial regulatory roles under salt and drought stress.

Conclusions: In summary, these results comprehensively describes the *WRKY* gene of five legumes, which provides a theoretical basis for further exploring the biological functions of members of the *WRKY* gene family in legumes.

Background

WRKY is named because of its special heptapeptide conserved sequence *WRKYGOK*, its one of the largest transcription factors family in plants [1]. Ishiguro et al [2] found the first plant *WRKY* gene in *Ipomoea batatas* and named it *Sweet Potato Factor1* (SPF1). Zhao et al [3] proposed that *WRKY* domain is mainly composed of a highly conserved *WRKYGQK* heptapeptide at the N-terminal and a zinc-finger motif (C_2H_2 or C_2HC) at the C-terminal. According to the number of *WRKY* domains and the type of the zinc-finger motif, the *WRKY* transcription factors are classified into three groups (I, II and III): group I members normally contain 2 *WRKY* domains; group II and III members all contain 1 *WRKY* domains and 1 zinc-finger motif (C_2H_2 or C_2HC) [1]. Most plant *WRKY* transcription factor has more group II members, and they are subdivided into five subgroups according to other conserved structural motifs: IIa IIb IIc IId IIe [4]. Group IIa *WRKY* genes widely participate in regulating various physiological process; Group IIc *WRKY* genes can resistance to salt stress by regulation of ABA signaling synthesis; Group IId *WRKY* genes can interact with the Ga^{2+} protein to help the plants defend against salt stress [5, 6, 7].

WRKY gene widely involved in regulating plant development and abiotic stress [1]. Pillai et al [8] found that the *OsWRKY42* transgenic strain had a significantly higher average weight (calculated by total fresh weight) than wild-type plants ($p < 0.05$) after stress treatment, proving that *OsWRKY42* enhanced tolerance in transgenic plants. Jiang et al [9] a *WRKY* gene transcription factor gene *PaWRKY77* was isolated from poplar trees. *PaWRKY77* reduces salt tolerance in transgenic poplars by binding to the W-box in the promoter and targeting ABA and salt-induced *PaNAC002* and *PaRD26* genes.

Most WRKY transcription factors function by binding to specific promoter. Previous study showed that *CsNIP* may repress the expression of other *AQPs* and induce osmoregulatory material accumulation to regulate water loss. Moreover, two WRKY, *CsWRKY4* and *CsWRKY28* were found to directly bind to the *CsNIP* promoter [10]. Lim et al [11] found that *OsWRKY5* directly binds to the W-box element of *OsMYB2* promoter to inhibit *OsMYB2* expression, thus downregulating the *OsMYB2* downstream genes in the ABA signaling pathway, indicating loss of *OsWRKY5* function might improve drought resistance in rice. The WRKY protein can also interact with its own family proteins or with other family proteins [12, 13]. Rice *OsWRKY51* can interact with *OsWRKY71* and participate in regulating the expression of amylase gene *Amy32b* [12]. Meanwhile, taking advantage of yeast two-hybrid technique, *AtWRKY10* was shown to interact with *AtVQ14* and regulate endosperm growth and seed development [14]. Moreover, WRKY acts antagonistically with VQ, chromatin immunoprecipitation analysis additionally shows that WRKY8 directly binds to the promoter of RD29A under salt conditions. These results indicated VQ9 protein acts as a repressor of WRKY8 factor to maintain the balance of WRKY8-mediated signaling path, thereby establishing salinity stress tolerance [15].

Legumes are one of the largest family in angiosperms, with soybean (*Glycine max*) growing most widely. Legumes are also one of the important sources of starch, oil and vegetables for human food, and are an important cash crop and feed crop in the world [16]. With the development of sequencing technology, more and more gene family are widely identified in different species, a total of 109, 72, 176, and 171 WRKY proteins were found in *Oryza sativa*, *Arabidopsis thaliana*, *Glycine max* and *Triticum aestivum*, respectively [17, 18, 19, 20]. However, the WRKY family members of multiple legumes have not been analyzed and studied so far. Here, we conducted a detailed bioinformatics analysis of the WRKY genes of five legume species, including gene identification, gene structure, gene duplication, chromosomal localization, protein interaction networks and miRNA prediction. These results provide theoretical support for further insights into the biological functions of WRKY transcription factors.

Materials And Methods

Identification of *WRKY* genes

Protein data and gene sequence information for five legumes were obtained from the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>) (Table S1). Download the sequence of members of the *A.thaliana* WRKY family on the PlantTFDB database website (<http://planttfdb.gao-lab.org/>). Identification of the WRKY gene using the blast search and hidden Markov model (HMM). BlistP and BlastN searches were performed using known *A.thaliana* and rice WRKY sequences, and preliminary identification of candidate WRKY genes was made. The conserved WRKY domain in each candidate WRKY protein was confirmed using the HMM atlas (login number PF03634) from the Pfam protein family database (<http://pfam.sanger.ac.uk>) [21]. The molecular weight (MW), isoelectric point (pI) and length of each WRKY protein were predicted by the online ExPASy programme (<http://www.expasy.org/tools/>) [22]. Detailed information on the WRKY protein in these five legumes is shown in the Table S2.

Phylogenetic analysis

Claustal X 2.1 was used to compare the WRKY gene sequences of five legumes to study the phylogenetic relationships of their WRKY gene families [23]. The MEGA4 program was used to build a neighbor connection (NJ) phylogeny tree [24]. Depending on the homologous distance or *A.thaliana*, *P.vulgaris*, *C.arietinum*, *M.runcatula*, *L.japonicus* and *G.max*.

Multiple alignment, conserved motif, and gene structure analysis of *WRKY* genes

Using the CD-search tool (<https://www.ncbi.nlm.nih.gov/cdd/>) in NCBI to predict the conservative domains of genes, we used Batches CD-Search to submit sequences to search for conservative domains. The genetic structure of the WRKY family, such as exons and introns, was extracted from the genome annotation information of five leguminous crops, and submitted to the website tool GSDS for analysis and display. Using the MEME online program, the matrix of WRKY proteins of five legumes was predicted and analyzed, the predicted quantity was 15, the length of the motif was from 20 to 300, and the number of motif replicates was "0 or 1", and other parameters were defaulted, the predicted results were sent to the mailbox, and the XML files of mast and meme were downloaded. Finally, the software TBtools were used to visually analyze the phylogenetic tree, motif, intron exons. Submit the downloaded genome annotation file to the software TBtools for visual analysis of the gene structure.

Gene duplication and collinearity analysis

Circos was used to locate the physical location of the WRKY gene on the chromosome. Synchronicity analysis between legume genomes was performed locally, using a similar type to PGDD (<http://chibba.agtec.uga.edu/duplication/>) [25]. We first used BLASTP and OrthoMCL software (<http://orthomcl.org/orthomcl/about.do#release>) to search for potential homologous gene pairs in multiple genomes ($E < 1e-5$, top 5 matched pairs). These homologous pairs are then entered into the PGDD database (<http://chibba.agtec.uga.edu/duplication/>) to identify collinear gene strands. Interspecific collinear analysis of the genomes of five leguminous crops was performed using MCScanX technology. Using TBtools, the collinearity of members of the *WRKY* gene family between the genomic species of five legume crops was drawn.

Ka and Ks calculations for *WRKY* genes

Ka/Ks could be used to evaluate selection history or divergence time [26]. We used the KaKs Calculator 2.0 and NG methods to calculate the number of non-synonymous (Ka) and synonymous (Ks) substitutions for the WRKY gene [27].

Cis-elements analysis of the *WRKY* promoter

Analysis of cis-elements regulatory of the WRKY promoter to further understand the WRKY transcription factor. We extracted from a genome-wide database of five leguminous crops, extracted the WRKY gene of each species, the promoter cis-regulator elements in the 1500 bp region upstream of the starting codon,

and then used the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) to search for cis-elements of promoters, and analyze the type, number and function of cis-acting elements. Finally, the cis-acting elements in the promoter were visualized by TBtools software.

WRKY protein interaction network and target genes network for miR5658 and miR5021

We utilized the miRBase database (<http://www.mirbase.org/>) to find mature mi5658 and miR5021 sequences. We analyzed the coding sequence of WRKY using psRNAT (<http://plantgrn.noble.org/psRNATarget>) and predicted the target of miR5021. To explore the sequence of WRKY proteins, we selected protein sequences of 50 WRKY transcription factors in the genomic database of several legumes and localized them on AtWRKY proteins in The Arabidopsis Information Resource database. Afterwards, We used the PAIR website (<http://www.cls.zju.edu.cn/pair/>) to predict the interaction between WRKYs and other proteins, and plotted the network of interactions in Cytoscape 3.3.0.

Plant material and treatments

Soybean seeds (William82) were used in this study. The seed was donated by the Heilongjiang Academy of Agricultural Sciences. We planted soybean seeds in a mixture of soil and sand ratios of 3:1 (w/w), germinated under greenhouse conditions, and irrigated every 2 days with 1/2 intensity Hoagland solution [28]. Seedlings grow at nighttime temperature 18 °C, daytime temperature 24 °C, relative humidity 67-70%, photoperiod 14/10h (six o'clock in the morning to eight o'clock in the evening), light intensity 180-200 $\mu\text{mol m}^{-2} \text{s}^{-2}$. After 8 weeks, treat the sprouted seedlings with 300 mM mannitol (drought) or 200 mM sodium chloride solution (salt). The whole seedlings were harvested 1h, 6h and 12 h after the control group and the treatment group, respectively. Samples after treatment should be frozen immediately in liquid nitro and stored at -80 °C until RNA extraction is used upon removal.

Synthesis of cDNA, RNA extraction, and real-time PCR fluorescence quantification

Experimental procedures conform to MIQE guidelines [29]. According to the manufacturer's agreement, a pure plant kit prepared using RNA (Tiangen, Beijing, China) extracts total RNA from whole seedlings. Stain with ethidium bromide (EB) using a 1.0% (w/v) agarose gel to assess RNA quality and then remove DNA contamination with DNaseI treatment (Takara, Shiga-ken, Japan). We synthesized cDNA using the Transcription First Strand cDNA Synthesis Kit (Indianapolis, IN, USA) and used it as a template for gene expression analysis. We needed to store the resulting cDNA at - 80 °C. For better research, qRT-PCR in our experiments was conducted using SYBR Green (TransGen Biotech, Beijing, China) and monitored on an ABI7500 real time PCR system (Applied Biosystems, Foster City, CA, USA). In order to analyze the expression of WRKY gene, the analysis method of DNA and qRT-PCR melt curve was used. Using β -actin as internal control of the *G.max*, the $2^{-\Delta\Delta C_t}$ method was used to normalize gene expression. Our primer design references WRKY CDS (*G.max*), Primer Express 3.0 software [30]. All primers used by qRT-PCR are shown in Table S3.

Results

Protein analysis and identification of WRKY gene in legumes

A chromosomal localization map was constructed to depict the localization of WRKY gene in every chromosome. 527 WRKY genes identified by HMMER analysis and Blast binding, 90 from *P.vulgaris*, 83 from *L.corniculatus*, 61 from *C.arietinum*, 108 from *M.truncatula* and 185 from *G.max*, respectively. WRKY family members of five legumes (*Phaseolus vulgaris*, *Medicago truncatula*, *Lotus japonicus*, *Glycine max*, *Cicer arietinum*) were identified based on genome-wide protein sequence information. Scalability predicts that 527 WRKY proteins have distinct physical and chemical properties, their isoelectric points range from 4.62 (*LjWRKY31*) to 10.11 (*CaWRKY41*), and the molecular weights range from 6413.2 (*MtWRKY43*) to 155716.32 (*GmWRKY46*) kDa. Information of the WRKY proteins is shown in Table S2.

Phylogenetic analysis of the WRKY genes in legumes

To further investigate the phylogenetic relationships of the WRKY gene family in five legumes and *A. thaliana*, the WRKY proteins in all six species were divided into three categories and five subclasses (I, IIa, IIb, IIc, IId, IIe and III). One hundred and twenty-eight WRKY genes belong to class I, 456 WRKY genes belong to class II, and 15 WRKY genes belong to class III.

In the five legumes, the I, IIa, IId, IIe, and III members were found to be embedded in other groups or subgroups, and had the largest proportion of genes belong to class II genemembers, *G.max*77.30% (143/185), *C.arietinum*77.05%(47/61), *P.vulgaris*76.67%(69/90), *M.truncatula*70.37% (76/108) and *L. japonicus*79.52% (66/83); Class I gene family has the second largest number of genes in addition to class II gene families, *G. max*21.62%(40/185), *C. arietinum* 22.95%(14/61), *P. vulgaris* 22.22%(20/90), *M. truncatula* 19.44% (21/108) and *L.japonicus*18.07% (15/83). The smallest proportion is class III gene family, with the amount of *G.max*1.08% (2/185), *C.arietinum*0%(0/61), *P.vulgaris* 1.11%(1/90), *M.truncatula*10.19% (11/108) and *L.japonicus*2.41% (2/83) (Fig. 1).

Conserved domains, gene structures, and motif analysis

A total of 15 motifs (motif1-motif15) were identified in all *WRKY* gene family members. Motif1 and motif2 are present in almost all of *WRKY* genes, with class I members having motif1/3/6/13/15, class II members having all 15 motifs, and class III members having motif1/3/7/11/13/15 (Fig. 2). Compared to other subfamilies, most class I family members have shorter WRKY protein sequences and relatively more conserved motif composition.

While we studied the conserved domain, we found that the different functions that exon-introns of the WRKY gene are also important. Gene structure analysis showed that WRKY in all five legume species consists of different number of exon-intron. The exon-intron of Class I, II, and III gene families are 2-4, 2-8 and 2-3, respectively. There are 88 out of the 63 genes contain only one intron, and 527 genes do not

contain introns. The intronic regions of subgroup II members are longer than the other subgroups, with good structural diversity and are not conserved (Fig. 2).

Chromosome mapping and gene duplication analysis of five legumes

The five legume WRKY genes are unequally distributed across the chromosomes. For *Phaseolus vulgaris*, *Medicago truncatula*, *Lotus japonicus*, *Glycine max* and *Cicer arietinum*, WRKY genes are distributed on 11, 8, 6, 20 and 7 chromosomes, respectively. The Gm6 chromosome contains the largest number of WRKY genes, and the Pv11 chromosome contains the least number of WRKY genes. According to Holub's definition of Gene cluster, gene replication events often occur during the evolution of organisms[31]. Although gene replication are frequent in the WRKY gene family, they form only a small number of gene clusters. In this study, 10, 15, 3, 22 and 5 gene clusters including 130 genes were found in the WRKY genes of *Phaseolus vulgaris*, *Medicago truncatula*, *Lotus japonicus*, *Glycine max*, *Cicer arietinum* respectively(Fig. 3).

A total of 116 gene replication events were detected in the WRKY gene of five legumes(*Phaseolus vulgaris*10, *Medicago truncatula*15, *Lotus japonicus*5, *Glycine max*81, *Cicer arietinum*5), involving 195 genes, covering 37% of all WRKY genes. Besides, 122 genes belong to the segmental duplications and 130 genes belong to tandem duplications, which suggested that the formation of these WRKY genes in five legume species may not depend primarily on gene duplication(Fig. 3). In order to obtain the Ka-Ks ratio, we calculated the Ka/Ks value of the gene pair with replication relationships, we found that the Ka, Ks ratios of these legumes were distributed between 0 and 0.5, which is less than 1, the suggests that these gene pairs have undergone strong purification selections.

Evolutionary analysis of WRKY genes in multiple legumes

To further study the phylogenetic mechanism of the WRKY family in legumes, we performed collinear analysis of WRKY genes in five dicots. The speculation results show *M. truncatula*/*C.arietinum* (117 pairs), *C.arietinum*/*G.max* (338 pairs), *P.vulgaris*/*G. max*(954 pairs), and *L.japonicus*/*P.vulgaris*(114 pairs). *P.vulgaris* (homologous pairs are mainly situated on 9 chromosomes) and *G. max* (homologous pairs are mainly located on 14 chromosomes) have the highest homology. Followed by *C.arietinum* and *G.max* (in addition to chromosome 2, homologous gene pairs are mainly distributed on 6 chromosomes) with second highest homology(Fig. 4).

Cis-acting regulatory elements in the promoter of the WRKY gene

Promoter sequences were analyzed by using PlantCARE to identify cis-regulatory elements within 1,500 bp upstream of the WRKY start codon. In this study, a total of 327 genes were found involving in ABRE (ABA) reaction; 318 genes related to light response elements; 175 genes were showed contain gibberellin response elements; 85 genes with auxin response elements; 24 genes related to trauma response elements; 8 genes involved in endosperm-specific negative expression. In addition, 320 MYB binding sites might involve in drought-induction, light-responsive, and flavonoid biosynthesis gene regulation (Fig. S2).

Analysis of WRKY gene regulatory networks in various legumes

We found 13 genes containing miR5021 recognition sites and 15 genes containing miR5658 recognition sites in these five legumes. By comparing multiple miR5658 and miR5021 maturation sequences, it was found that the miR5658/5021-WRKY regulatory module was highly conserved in five legumes. There is a mismatch between miR5021 and miR5658 between the WRKYmRNAs targeted in this study, their core target sequences are conserved and play an important regulatory role in the *WRKY* transcription factor (Fig. 5).

Moreover, we found that the WRKY proteins in five legumes were completely localized to the WRKY protein and other proteins in *Arabidopsis*. To study the interactions between other proteins and the *WRKY* gene, we made predictions using the PIRA tool. These WRKY genes interact with different proteins, including mitogen-activated protein kinase (MKK), transcription factor (MYB), nucleoside diphosphate kinase (NDPK), 1-aminocyclopropane-1-carboxylate synthase (ACS) and other proteins. In addition, we also found interactions of other WRKY proteins, such as *WRKY25/WRKY48*(*MtWRKY55*), *WRKY33*(*LjWRKY20/37/43*)/*WRKY48*(*MtWRKY55*), and *WRKY72*(*GmWRKY78*)/*WRKY9* (*MtWRKY48/68*, *GmWRKY104*) (Fig. 5).

Specific expression of the *WRKY* gene

In the tissues of beans, alfalfa, soybean and chickpea, class I family members (*PvWRKY60* and *GmWRKY177*) are highly expressed in flowers, *MtWRKY49* is highly expressed in the roots, the expression level of *GmWRKY47/58/82* are higher in leaves, the remaining members were less expressed in each tissue. Class II family members (*PvWRKY4/29/38*) are highly expressed in 10cm length roots, (*PvWRKY4/29/83*) higher expressed in 10cm trunk. *PvWRKY66*, *MtWRKY50/82*, *GmWRKY88* are all highly expressed in flowers. *MtWRKY1/19/50*, *LjWRKY68*, *GmWRKY42/45/81/123* and *CaWRKY8/56* are all highly expressed in roots. *PvWRKY29*, *MtWRKY41/50* and *CaWRKY43* are all highly expressed in the root nodules. The expression level of *MtWRKY82* and *CmWRKY100* are high in leaves. Moreover, *MtWRKY82* is also highly expressed in the seedpod. The remaining members showed low expression in each tissue. For Class III family members, in addition to the high expression of *MtWRKY50* in flowers, the remaining members are barely expressed in each organism (Fig. 6).

Expression of WRKY genes under abiotic stress

The *WRKY* gene family plays critical roles in plant's abiotic stress. In this study, In order to study the expression of soybean WRKY gene under drought stress and salt stress, the qRT-PCR method was used. Our results indicate that many genes behave differently in different tissues. Under salt stress, 21 genes were upregulated during stress treatment, 5 genes (*GmWRKY39/75/130/132/160*) were highly expressed, 4 genes (*GmWRKY39/75/160*, *GmWRKY130*) expression peaked at 6h and 12h, respectively. 14 genes were continuously upregulated at 1h, 6h and 12h (Fig. 7a). Under drought stress, most genes were down-regulated at all treatment times, only 1 gene (*GmWRKY29*) upregulated and its expression peaked at 6h. Besides, (Fig. 7b)

Discussion

WRKY transcription factors have been found in many plants, such as *Arabidopsis thaliana* [18], rice [17], wheat [20] and barley [32]. It was shown that WRKY can affect plant growth and development, as well as response to biotic and abiotic stresses [33]. However, previous studies were only based on evolutionary analysis of WRKY gene family composition in individual species, rarely compared WRKY gene families among different species. In this study, 185, 61, 90, 108 and 83 genes containing the WRKY motif were identified in *M. truncatula*, *P. vulgaris*, *C. arietinum*, *L. japonicus* and *G. max*, respectively. The WRKY gene of five legumes can be divided into 7 subfamily: I, IIa, IIb, IIc, IId, IIe and III, which is similar to the classifications of WRKY genes in *Arabidopsis* and potato [34, 35]. Moreover, class II has the largest number of members, followed by class I, class III and low diversity, consistent with the results from Wu et al [5].

Although the domain of the WRKY gene family is highly conserved in legumes, substitutions occur in its amino acid sequences. For example, *WRKYGKK* in maize and soybean [1]. In *Oryza sativa*, members of the WRKY family have 19 variants of the WRKY domain, *WRKYGKK* and *WRKYGEK* are two common variants with 5 and 7 domains, respectively; And other variants (including *WKKYGQK*, *WRICGQK*, *WRKYSEK*, *WKRYGQK*, *WSKYEQK*, *WRMCGQK* and *WIKYGQK*) [36]. WRKY genes with similar genetic structure and motif composition were clustered together by phylogenetic analysis. The *CaWRKY* gene always comes together with the *MtWRKY* gene, suggesting that they are evolutionarily more closely related. Moreover, the same exists were found between *GmWRKY* and *PvWRKY*, and between *MtWRKY* and *AtWRKY*.

Our gene replication analysis of the WRKY gene family showed that segmental replication had less effect on WRKY gene amplification than fragment replication in five legumes, which are similar to the observation in peanut and purple [37]. Chanwala et al [38] studied WRKY genes in multiple monocot crops and found that WRKY genes may be unevenly distributed across the genome and chromosomes, which was also found in legume dicot crops, therefore we speculate that this phenomenon might related to segmental duplication. Wang et al [39] proposed that soybean is diploid and has undergone at least one genome-wide doubling event during evolution. In this study, we found that soybean had the most gene family members and fragment duplication events compared to the other four legumes. Thus, We speculate that this may due to the diploidization of the soybean during its evolution.

Protein interaction network analysis revealed that *WRKY33* homologs *LjWRKY20/37/43* all contain *MYB* binding sites and can interact with *MYB51* protein. Tao et al [40] found that *WRKY33* can induce the expression of *MYB51* and *CYP83B1* and stimulate the biosynthesis of the precursor *I3G* of *4MI3G*, thus inhibiting the formation of *S. rape*. MPK3 protein, which is considered one of the important components of the model plant clock. Sheikh et al [41] found that MPK3 can phosphorylate *WRKY46*, leading to the activation of WRKY transcription factor, thus actively regulating the expression of defense genes. There is a protein interaction network in which the six WRKY proteins interact with the MPK3 protein. We speculate that MPK3 may phosphorylate these six WRKY transcription factors. In addition, different WRKY

members have different hormone-related cis-acting elements, such as TGA element (auxin)/ABRE (ABA), P-box (gibberellin)/GARE motif and WRKY element (Figure S2). The results suggested that the *WRKY* gene may also play an important role in the regulatory pathways of plant signaling.

In this study, *GmWRKY42/45/81/123* genes were highly expressed in roots. Zhou et al [42] confirmed that these genes were indeed highly expressed in roots by RT-PCR experiments. Besides, Wang et al [43] found that these genes were also overexpressed in transgenic roots. *PvWRKY4* was highly expressed in all tissues, consistent with the study of Wu et al [44]. Markus et al [45] found that *WRKY6*, an *ODAR* gene of *PvWRKY4*, also regulates root development. Wu et al [44] found one gene, *Pvwrky35*, is highly expressed in nodules and roots. Additionally, in this study, we found *PvWRKY29* is also highly expressed in nodules and roots. With transgenic methods, Khandal et al [46] found that *CaWRKY31* of *PvWRKY29* drives the expression of *CaCKX6* in root expansion and CKX activity in T4 transgenic chickpea plants.

It's well known that WRKY are involved in plant development and growth, and they can respond to multiple abiotic stresses. In this study, *GmWRKY13* was shown upregulated under drought stress indicating their drought-tolerant ability. Zhou et al [42] used transgenic methods to further investigate the function of *GmWRKY13* stress-induced genes and found that the reason of increased tolerance of *GmWRKY13* was due to its regulation of *DREB2A* and *STZ / Zat10*. Furthermore, most WRKY genes were downregulated under drought and salt stress, indicating that each WRKY member may share a common function under drought and salt stress response. Taken together, these results suggest that WRKY genes play an important role in responding to abiotic stress in plants.

Conclusions

A total of 527 *WRKY* genes were identified and analyzed comprehensively in this study. The chromosomal location, evolutionary relationships, conserved motifs, and gene structure of the *WRKY* genes were investigated in five legume species. The expression patterns of *WRKY* genes in seven different tissues and their responses to various stresses suggest that they may play important roles in legume growth and development.

Declarations

Ethics approval and consent to participate

All experimental studies, including the collection of plant materials, are permitted by Daqing Normal College in Heilongjiang Province and are in accordance with relevant institutional, national and international norms and legislation.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available in (<https://www.peanutbase.org/blast>).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the grant from the Daqing Normal University Doctoral Research Fund China (No. 21ZR05) and the Heilongjiang Natural Science Foundation Joint Guiding Project (LH2020C070).

Authors' Contributions

L.L. and N.C. conceived and designed the experiments. L.L. and H.Z. wrote the manuscript. L.L. and S.Q. conducted experiments and analyzed data. X.Z., Y.Y., Y.H. and G.R. performed the qRT-PCR experiments. L.L. and H.Z. revised the manuscript.

Acknowledgements

We thank all our colleagues and students in our lab for their useful discussions and technical assistance.

References

1. Thomas Eulgem,,Paul J. Rushton,,Silke Robatzek & Imre E. Somssich.(2000).The WRKY superfamily of plant transcription factors. Trends in Plant Science(5), doi:10.1016/S1360-1385(00)01600-9.
2. Sumie Ishiguro & Kenzo Nakamura.(1994).Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and β -amylase from sweet potato. MGG Molecular & General Genetics(6), doi:10.1007/BF00282746.
3. Zhao Nannan,,He Meijing,,Li Li,... Yang Xinlei.(2020).Identification and expression analysis of WRKY gene family under drought stress in peanut (*Arachis hypogaea* L.)..PloS one(4), doi:10.1371/journal.pone.0231396.
4. Y. GUO,,Z. CAI & S. GAN.(2004).Transcriptome of Arabidopsis leaf senescence. Plant, Cell & Environment(5), doi:10.1111/j.1365-3040.2003.01158.x.
5. Wu, J., Chen, J., Wang, L., & Wang, S. (2017). Genome-wide investigation of WRKY transcription factors involved in terminal drought stress response in common bean. Frontiers in Plant Science, 8, 380.
6. Park, C. Y., Lee, J. H., Yoo, J. H., Moon, B. C., Choi, M. S., Kang, Y. H., ... Cho,M. J. (2005). WRKY group IId transcription factors interact with calmodulin. FEBS letters, 579(6), 1545–1550.

7. Karim, A., Jiang, Y., Guo, L., Ling, Z., Ye, S., Duan, Y., ... Luo, K. (2015). Isolation and characterization of a subgroup IIa WRKY transcription factor PtrWRKY40 from *Populus trichocarpa*. *Tree Physiology*, 35(10), 1129–1139.
8. Pillai Shakuntala E., Kumar Chandan., Patel Hitendra K & Sonti Ramesh V. (2018). Overexpression of a cell wall damage induced transcription factor, OsWRKY42, leads to enhanced callose deposition and tolerance to salt stress but does not enhance tolerance to bacterial infection.. *BMC plant biology*(1), doi:10.1186/s12870-018-1391-5.
9. Jiang Yuanzhong, Tong Shaofei, Chen Ningning, Liu Bao, Bai Qiuxian, Chen Yang... Liu Huanhuan. (2020). The PalWRKY77 transcription factor negatively regulates salt tolerance and ABA signaling in *Populus*.. *The Plant journal: for cell and molecular biology*(), doi:10.1111/TPJ.15109.
10. Zhang Mingfei, Zhu Yanfei, Yang Hongbin, Li Xin, Xu Rangwei, Zhu Feng & Cheng Yunjiang. (2021). CsNIP5;1 acts as a multifunctional regulator to confer water loss tolerance in citrus fruit. *Plant Science*(prepublish), doi:10.1016/J.PLANTSCI.2021.111150.
11. Lim Chaemyeong, Kang Kiyoon, Shim Yejin, Yoo SooCheul & Paek NamChon. (2021). Inactivating transcription factor OsWRKY5 enhances drought tolerance through abscisic acid signaling pathways.. *Plant physiology*(), doi:10.1093/PLPHYS/KIAB492.
12. Xie Zhen., Zhang Zhong-Lin., Zou Xiaolu, ... Shen Qingxi J. (2006). Interactions of two abscisic-acid induced WRKY genes in repressing gibberellin signaling in aleurone cells.. *The Plant journal: for cell and molecular biology*(2), doi:10.1111/j.1365-313X.2006.02694.x
13. Ye Yu-Jie., Xiao Yun-Yi., Han Yan-Chao, ... Chen Jian-Ye. (2016). Banana fruit VQ motif-containing protein5 represses cold-responsive transcription factor MaWRKY26 involved in the regulation of JA biosynthetic genes.. *Scientific reports*(1), doi:10.1038/srep23632.
14. Cheng Yuan., Zhou Yuan., Yang Yan, ... Chen Zhixiang. (2012). Structural and functional analysis of VQ motif-containing proteins in *Arabidopsis* as interacting proteins of WRKY transcription factors.. *Plant physiology*(2), doi:10.1104/pp.112.196816.
15. Hu Yanru., Chen Ligang., Wang Houping, ... Yu Diqu. (2013). *Arabidopsis* transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance.. *The Plant journal: for cell and molecular biology*(5), doi:10.1111/tpj.12159.
16. Ohri, P., Bhardwaj, R., Kaur, R., Jasrotia, S., Parihar, R. D., & Sharma, N.. (2021). 14 - MicroRNAs and abiotic stress tolerance in legumes. Elsevier Inc.
17. Christian A. Ross., Yue Liu & Qingxi J. Shen. (2007). The WRKY Gene Family in Rice (*Oryza sativa*). *Journal of Integrative Plant Biology*(6), doi:10.1111/j.1744-7909.2007.00504.x.
18. Qishan Wang., Minghui Wang., Xiangzhe Zhang, ... Yuchun Pan. (2011). WRKY gene family evolution in *Arabidopsis thaliana*. *Genetica*(8), doi:10.1007/s10709-011-9599-4.
19. Song, H., Wang, P., Hou, L., Zhao, S., & Wang, X.. (2016). Global analysis of wrky genes and their response to dehydration and salt stress in soybean. *Frontiers in Plant Science*, 7, 9.
20. Pan Ning., Congcong Liu., Jingquan Kang & Jinyin Lv. (2017). Genome-wide analysis of WRKY transcription factors in wheat (*Triticum aestivum* L.) and differential expression under water deficit

- condition. PeerJ(), doi:10.7717/peerj.3232.
21. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J (2014) Pfam: the protein families database. *Nucleic Acids Res* 42:222–230.
 22. Lei Ling, Wenrui Zhang, Yimin An... Changhong Guo. (2020). Genome-wide analysis of the TCP transcription factor genes in five legume genomes and their response to salt and drought stresses. *Functional & Integrative Genomics*(4). doi:10.1007/s10142-020-00733-0.
 23. Li Donghua, Liu Pan, Yu Jingyin... Zhang Xiurong. (2017). Genome-wide analysis of WRKY gene family in the sesame genome and identification of the WRKY genes involved in responses to abiotic stresses.. *BMC plant biology*(1). doi:10.1186/s12870-017-1099-y.
 24. Khezri, G., Shobbar, Z. S., & Naji, A. M. (2015). In Silico Analysis of the WRKY Transcription Factors Gene Family in Wheat.
 25. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.
 26. Li, W. H., Gojobori, T., & Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. *Nature*, 292(5820), 237–239.
 27. Xu L, Qiao X, Zhang M, Zhang S (2018) Genome-wide analysis of aluminum-activated malate transporter family genes in six Rosaceae species, and expression analysis and functional characterization on malate accumulation in Chinese white pear. *Plant Sci* 24:431–442.
 28. Hoagland DR, Arnon DI (1950) The water culture method for growing plants without soil. *Calif Agric Exp Stn Circ* 347:1–38.
 29. Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M.,... Wittwer, C. T. (2009). The MIQE Guidelines: M inimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments.
 30. Junyi Lin, Yulei Yang, Ziqin Zhao... Mingchang Zhang. (2019). Mutations of Desmoglein-2 in Sudden unexplained death in the chinese han population. *Journal of Forensic Science and Medicine*(2). doi:10.4103/jfsm.jfsm_40_18.
 31. Holub, E. B.. (2001). The arms race is ancient history in arabidopsis, the wildflower. *Nature Reviews Genetics*, 2(7), 516–527.
 32. Zheng Junjun, Zhang Ziling, Tong Tao, Fang Yunxia, Zhang Xian, Niu Chunyu... Zhang Xiaoqin. (2021). Genome-Wide Identification of WRKY Gene Family and Expression Analysis under Abiotic Stress in Barley. *Agronomy*(3), doi:10.3390/AGRONOMY11030521.
 33. Huang Kai, Wu Tao, Ma Ziming, Li Zhao, Chen Haoyuan, Zhang Mingxing... Du Xinglin. (2021). Rice Transcription Factor OsWRKY55 Is Involved in the Drought Response and Regulation of Plant Growth. *International Journal of Molecular Sciences*(9), doi:10.3390/IJMS22094337.
 34. Wei, X., Yao, W., Jiang, T., Zhou, B., & University, N. F.. (2016). Identification of wrky gene in response to abiotic stress from wrky transcription factor gene family of arabidopsis thaliana. *Journal of Northeast Forestry University*.

35. Liu, Q. N., Liu, Y., Xin, Z. Z., Zhang, D. Z., Ge, B. M., Yang, R. P., ... Zhou, C.L. (2017). Genome-wide identification and characterization of the WRKY gene family in potato (*Solanum tuberosum*). *Biochemical Systematics and Ecology*, 71, 212–218.
36. Zhang Yuanji & Wang Liangjiang.(2005).The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evolutionary Biology*(1), doi:10.1186/1471-2148-5-1.
37. Zhao Xuyao,Yang Jingjing,Li Gaojie,Sun Zuoliang,Hu Shiqi,Chen Yan... Hou Hongwei.(2021).Genome-wide identification and comparative analysis of the WRKY gene family in aquatic plants and their response to abiotic stresses in giant duckweed (*Spirodela polyrhiza*). *Genomics*(4),doi:10.1016/J.YGENO.2021.03.035.
38. Chanwala Jeky,,Satpati Suresh,,Dixit Anshuman,... Dey Nrisingha.(2020).Genome-wide identification and expression analysis of WRKY transcription factors in pearl millet (*Pennisetum glaucum*) under dehydration and salinity stress.. *BMC genomics*(1), doi:10.1186/s12864-020-6622-0.
39. Wang Longfei,Jia Guanghong,Jiang Xinyu,Cao Shuai,Chen Z Jeffrey & Song Qingxin.(2021).Altered chromatin architecture and gene expression during polyploidization and domestication of soybean.. *The Plant cell*(5), doi:10.1093/PLCELL/KOAB081.
40. Tao, H., Miao, H., Chen, L., Wang, M., Xia, C., Zeng, W., ... Wang, Q. (2021). WRKY33-mediated indole glucosinolate metabolic pathway confers host resistance against *Alternaria brassicicola*. *bioRxiv*.
41. Sheikh, A. H., Lennart, E. L., Pascal, P., Wolfgang, H., Sinha, A. K., & Dierk, S., et al. (2016). Regulation of wrky46 transcription factor function by mitogen-activated protein kinases in *Arabidopsis thaliana*. *Front Plant*, 7(R87), 61.
42. Zhou Qi-Yun,,Tian Ai-Guo,,Zou Hong-Feng,... Chen Shou-Yi.(2008).Soybean WRKY-type transcription factor genes, GmWRKY13, GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in transgenic *Arabidopsis* plants.. *Plant biotechnology journal*(5), doi:10.1111/j.1467-7652.2008.00336.x.
43. Wang, C., Deng, P., Chen, L., Wang, X., Ma, H., Hu, W., ... He, G. (2013). A wheat WRKY transcription factor TaWRKY10 confers tolerance to multiple abiotic stresses in transgenic tobacco. *PloS one*, 8(6), e65120.
44. Wu, M., Liu, H., Han, G., Cai, R., Pan, F., & Xiang, Y. (2017). A moso bamboo WRKY gene PeWRKY83 confers salinity tolerance in transgenic *Arabidopsis* plants. *Scientific reports*, 7(1), 1–16.
45. Markus G. Stetter,,Martin Benz & Uwe Ludewig.(2017).Increased root hair density by loss of WRKY6 in *Arabidopsis thaliana*. *PeerJ*() , doi:10.7717/peerj.2891.
46. Khandal, H., Gupta, S. K., Dwivedi, V., Mandal, D., Sharma, N. K., Vishwakarma, N.K., ... Chattopadhyay,D.(2020). Root-specific expression of chickpea cytokinin oxidase/dehydrogenase 6 leads to enhanced root growth, drought tolerance and yield without compromising nodulation. *Plant biotechnology journal*, 18(11), 2225–2240.

Figures

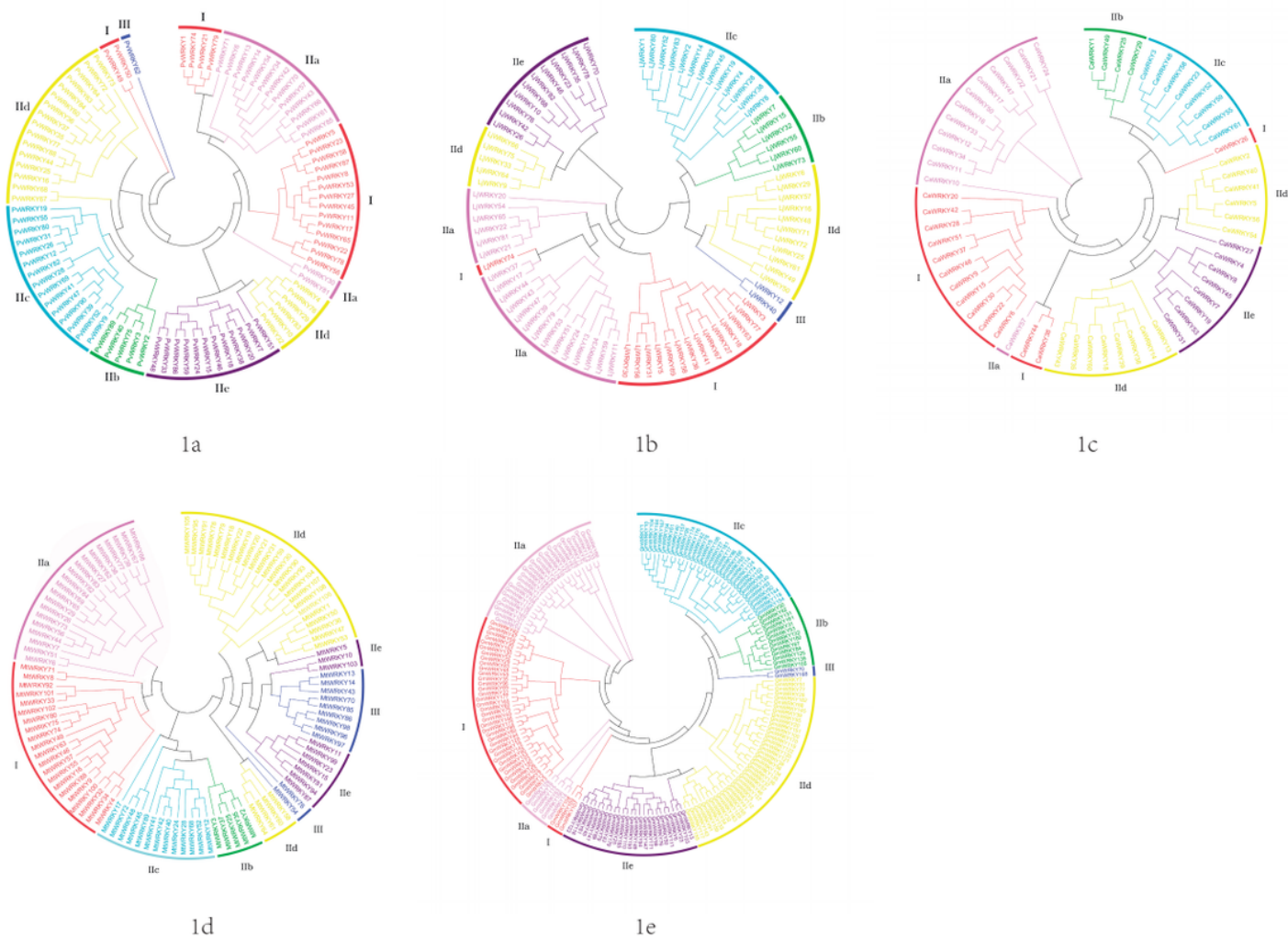


Figure 1

Phylogenetic relationships of the WRKY gene family of five legumes. Phylogenetic tree analysis of the WRKY genes in *Medicago truncatula*, *Cicer arietinum*, *Glycine max*, *Phaseolus vulgaris*, *Lotus japonicus*. The three clusters were designated as subfamily I-red, IIa-pink, IIb-green, IIc-ld-yellow, IIe-blue purple and III-navy blue.

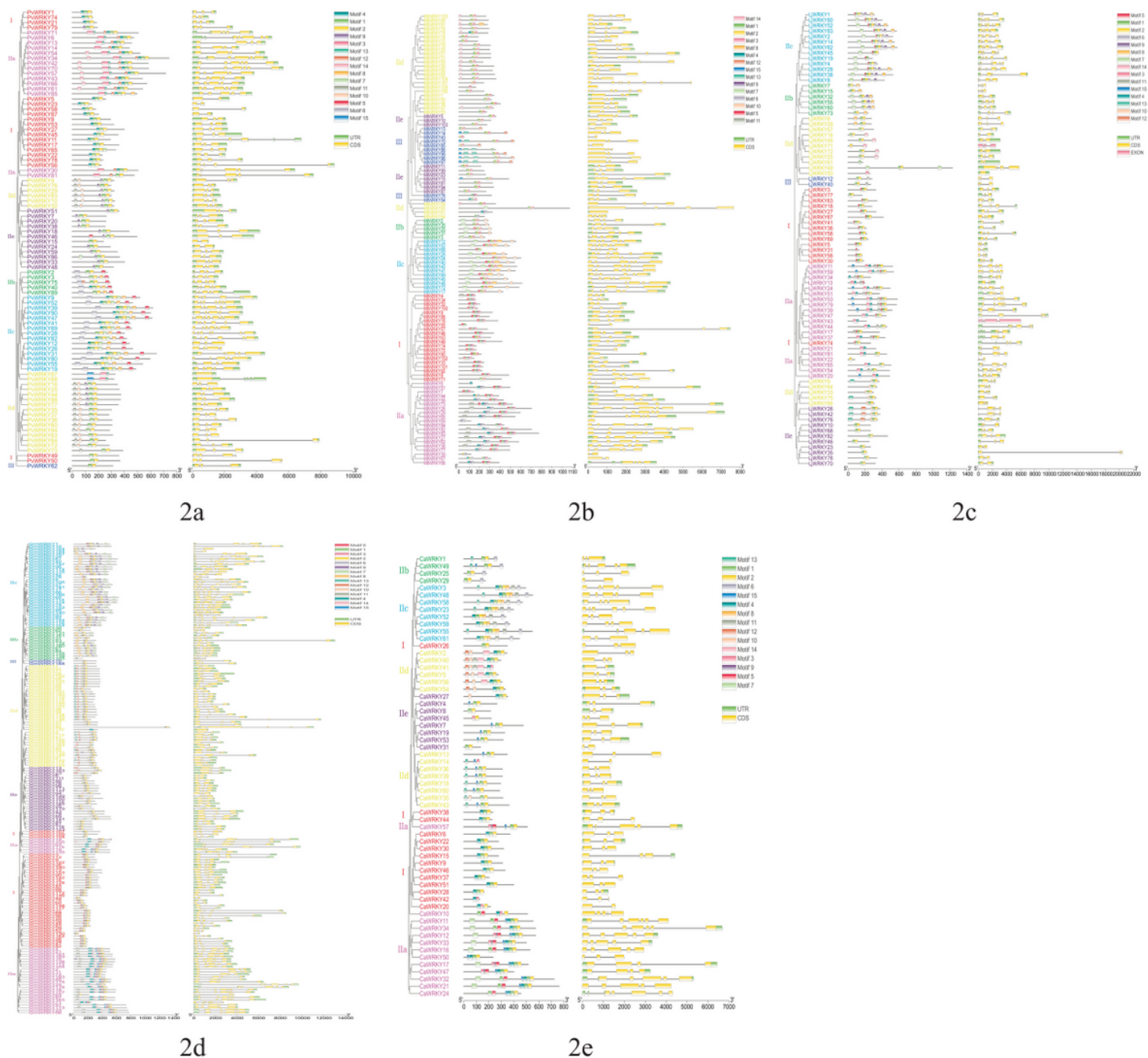


Figure 2

Motif and intron-exon plots of the WRKY gene family of five legumes. a *Phaseolus vulgaris*, b *Medicago truncatula*, c *Lotus japonicus*, d *Cicer arietinum* and e *Glycine max* WRKY gene family intron-exon and conserved domain analysis of the WRKY protein phylogenetic tree in the first column of each figure; The motif composition of the second column of the WRKY protein, Protional motif length; exon-intron structure of the third WRKY gene; green box-intron; yellow box-exon; pink box-non-coding region; Black line-5'end to 3'end non-coding region.

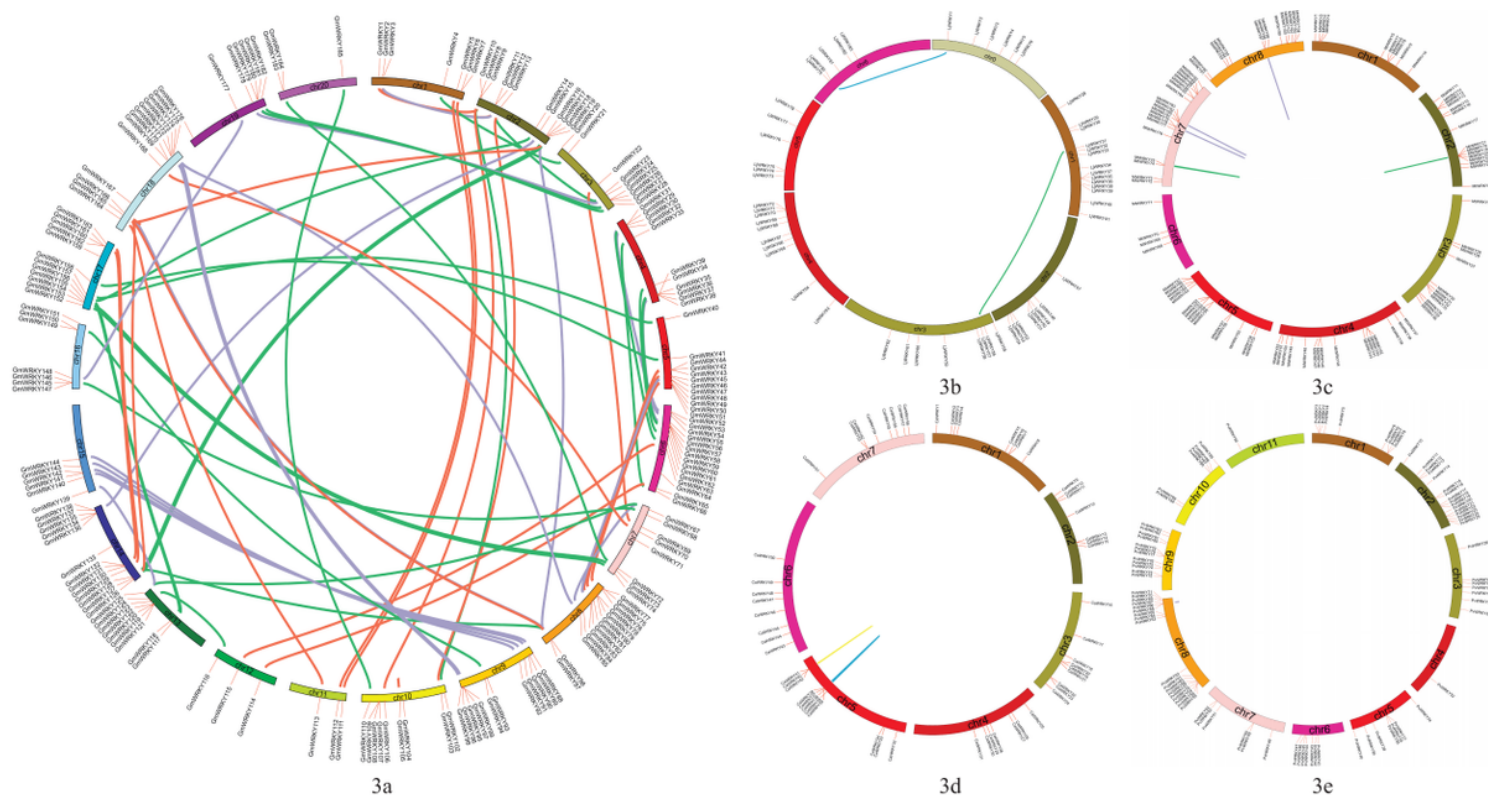


Figure 3

Gene reproduction of the WRKY gene family of five legumes. Chromosome location and duplication events analysis in a *Glycine max*, b *Lotus japonicus*, c *Medicago truncatula*, d *Cicer arietinum*, and e *Phaseolus vulgaris*. Genes of different subfamilies with gene duplication relationships were connected by lines in different colours.

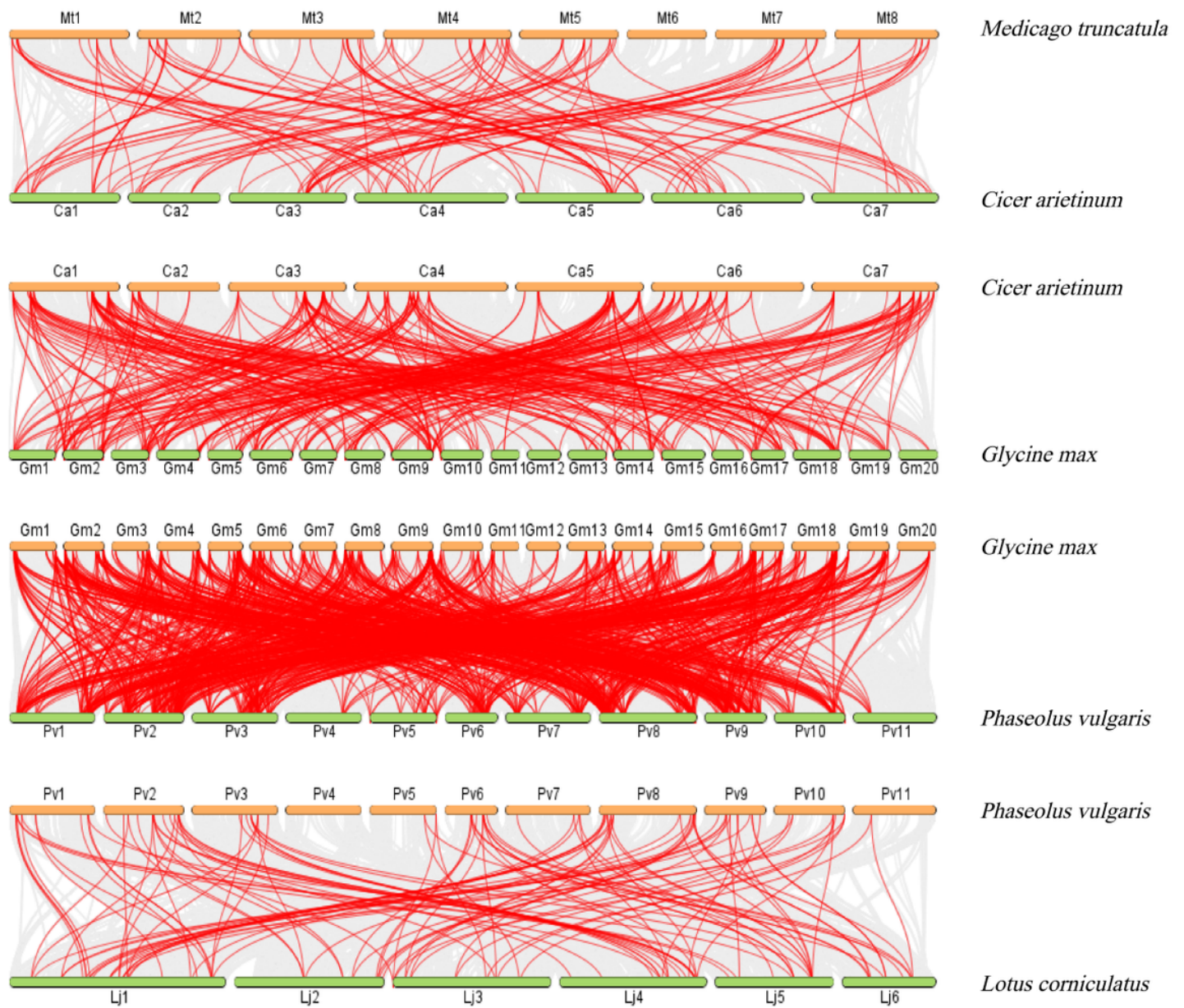
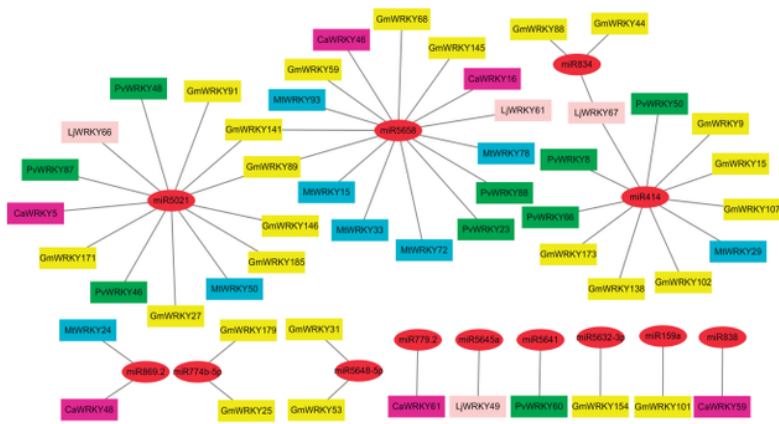
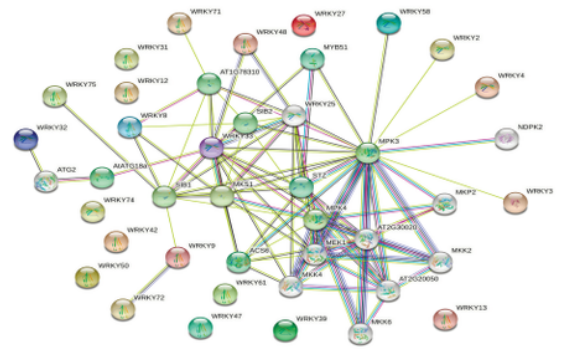


Figure 4

Synteny analysis of *WRKY* genes among five legume species.



5a



5b

Figure 5

The miRNAs targeted *WRKY* and gene regulatory network. a Aregulatory network relationships between the *WRKY* and other proteins, and the putative miRNAs and their targeted *WRKY* genes .b *WRKY* protein interaction network.

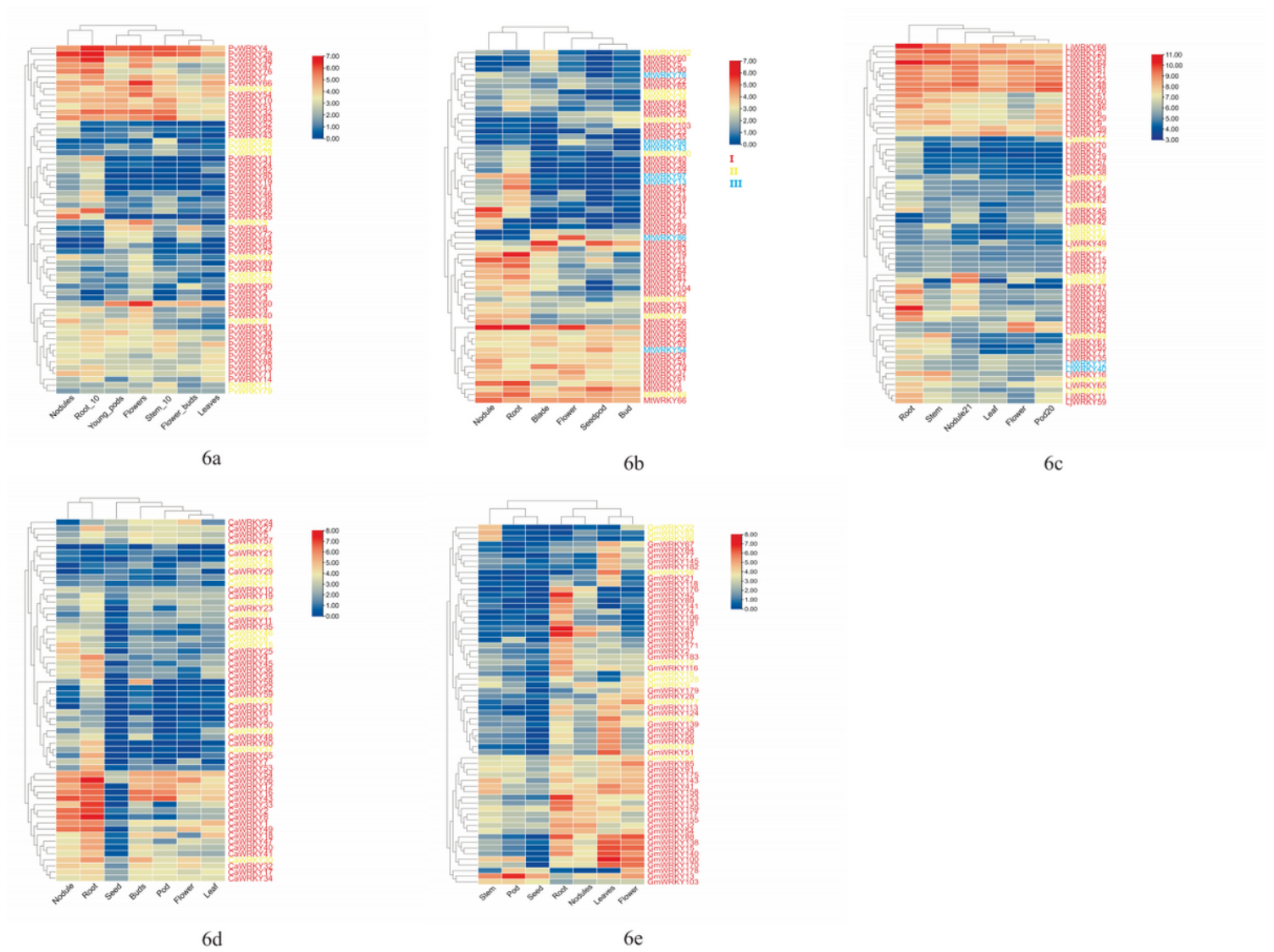


Figure 6

Expression analysis of WRKY gene in different tissues of five legumes. Expression analysis of *WRKY* genes in different tissues in a *Phaseolus vulgaris*, b *Medicago truncatula*, c *Lotus japonicus*, d *Cicer arietinum* and e *Glycine max*. Subgroup I-yellow, Subgroup II-red, Subgroup III-blue.

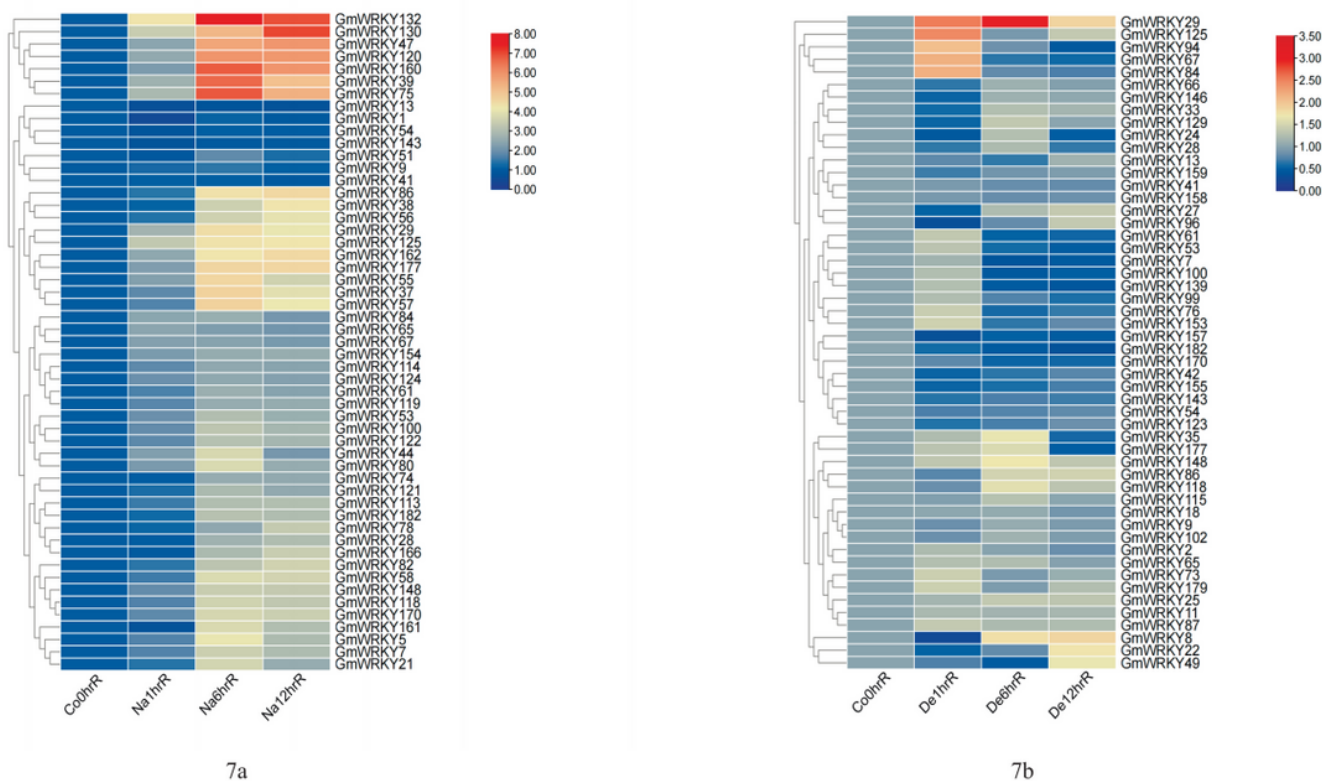


Figure 7

Expression analysis of the *WRKY* genes. qRT-PCR analysis reveals *WRKY* genes under a salt stress and b drought stress in *Glycine max*.

Supplementary Files

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

- [SupplementaryFigures.pdf](#)
- [TableS1DetailsofWRKYsequences.xlsx](#)
- [TableS2ListofallWRKYgenesinformationidentifiedinthefivelegumegenomes.xlsx](#)
- [TableS3PrimersequencesusedinqRTPCRanalysis.docx](#)