Genome-wide analysis of the WRKY gene family in drumstick (*Moringa oleifera* Lam.)

Junjie Zhang [Equal first author, 1, 2, 3, 4] , **Endian Yang** [Equal first author, 4] , **Qian He** [3, 4] , **Mengfei Lin** [1, 2, 3, 4] , **Wei Zhou** [1, 2, 3, 4] , **Ruiqi Pian** [1, 2, 3, 4] , **Xiaoyang Chen** [Corresp. 1, 2, 3, 4]

1 State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources (South China Agricultural University), Guangzhou, China

2 Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, Guangzhou, China

3 Guangdong Province Research Center of Woody Forage Engineering Technology, Guangzhou, China

4 College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China

Corresponding Author: Xiaoyang Chen
Email address: xychen@scau.edu.cn

WRKY proteins belong to one of the largest families of transcription factors. They have important functions in plant growth and development, signal transduction and stress responses. However, little information is available regarding the WRKY family in drumstick (*Moringa oleifera* Lam.). In the present study, we identified 54 *MoWRKY* genes in this species using genomic data. On the basis of structural features of the proteins they encode, the *MoWRKY* genes were classified into three main groups, with the second group being further divided into five subgroups. Phylogenetic trees constructed from the sequences of WRKY domains and overall amino acid compositions derived from drumstick and Arabidopsis were similar; the results indicated that the WRKY domain was the main evolutionary unit of *WRKY* genes. Gene structure and conserved motif analysis showed that genes with similar structures and proteins with similar motif compositions were usually clustered in the same class. Selective pressure analysis indicated that although neutral evolution and positive selection have happened in several *MoWRKY* genes, most have evolved under strong purifying selection. Moreover, different subgroups had evolved at different rates. The levels of expression of *MoWRKY* genes in response to five different abiotic stresses (salt, heat, drought, H_2O_2, cold) were evaluated by reverse transcription polymerase chain reaction (RT-PCR) and quantitative RT-PCR (qRT-PCR), with the results indicating that these genes had different expression levels and that some may be involved in abiotic stress responses. Our results will provide a foundation for cloning genes with specific functions for use in further research and applications.

1   Genome-wide Analysis of the WRKY Gene Family in Drumstick (*Moringa oleifera* Lam.)

2   Junjie Zhang [1, 2, 3, 4 †], Endian Yang [4 †], Qian He [3, 4], Mengfei Lin [1, 2, 3, 4], Wei Zhou [1, 2, 3, 4], Ruiqi Pian [1, 2, 3, 4], Xiaoyang

3   Chen [1, 2, 3, 4*]

4   [1] State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources (South China Agricultural

5   University), Guangzhou, China

6   [2] Guangdong Key Laboratory for Innovative Development and Utilization of Forest Plant Germplasm, Guangzhou,

7   China

8   [3] Guangdong Province Research Center of Woody Forage Engineering Technology, Guangzhou, China

9   [4] College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China

11   ***Correspondence:**

12   Dr. Xiaoyang Chen

13   xychen@scau.edu.cn

14   †These authors have contributed equally to this work.

30 **Abstract**

31 WRKY proteins belong to one of the largest families of transcription factors. They have important functions in plant

32 growth and development, signal transduction and stress responses. However, little information is available regarding

33 the WRKY family in drumstick (*Moringa oleifera* Lam.). In the present study, we identified 54 *MoWRKY* genes in

34 this species using genomic data. On the basis of structural features of the proteins they encode, the *MoWRKY* genes

35 were classified into three main groups, with the second group being further divided into five subgroups.

36 Phylogenetic trees constructed from the sequences of WRKY domains and overall amino acid compositions derived

37 from drumstick and Arabidopsis were similar; the results indicated that the WRKY domain was the main

38 evolutionary unit of *WRKY* genes. Gene structure and conserved motif analysis showed that genes with similar

39 structures and proteins with similar motif compositions were usually clustered in the same class. Selective pressure

40 analysis indicated that although neutral evolution and positive selection have happened in several *MoWRKY* genes,

41 most have evolved under strong purifying selection. Moreover, different subgroups had evolved at different rates.

42 The levels of expression of *MoWRKY* genes in response to five different abiotic stresses (salt, heat, drought, H_2O_2,

43 cold) were evaluated by reverse transcription polymerase chain reaction (RT-PCR) and quantitative RT-PCR (qRT-

44 PCR), with the results indicating that these genes had different expression levels and that some may be involved in

45 abiotic stress responses. Our results will provide a foundation for cloning genes with specific functions for use in

46 further research and applications.

47

48

49

50

51

52

53

54

55

56

Introduction

Transcription factors (TFs), which bind to specific DNA motifs, are important in regulating gene expression and

controlling various important biological processes (Smith and Matthews, 2016). Out of numerous families of TFs,

the WRKY gene family, named after a conserved WRKY domain, is one of the largest, and it is known to be

involved in a range of plant processes from germination to senescence (Chen et al., 2012; Eulgem et al., 2000;

Pandey and Somssich, 2009; Ulker and Somssich, 2004). WRKY genes were first identified in plant species

(Ishiguro and Nakamura, 1994) and originally thought to be plant-specific (Eulgem et al., 2000). However, in recent

years WRKY proteins have been identified in non-plant species, such as *Giardia lamblia*, *Dictyostelium discoideum*

and so on (Li et al., 2016; Zhang and Wang et al., 2005). The WRKY domain contains about 60 amino acid residues,

comprising a highly conserved short amino acid sequence, WRKYGQK, at the N-terminus and an adjacent C_2H_2 or

C_2HC zinc finger structure (Eulgem et al., 2000). Depending on the number of WRKY domains and the type of zinc

finger, the WRKY family can be divided into three main groups. Group I contains two WRKY domains and the

C_2H_2 zinc finger type. Group II contains one WRKY domain and a C_2H_2 type zinc finger motif; this group can be

further divided into five subgroups, IIa, IIb, IIc, IId and IIe. The WRKYs with a single WRKY domain and a C_2HC

zinc-finger structure belong to group III (Eulgem et al., 2000; Goel et al., 2016; Li et al., 2017).

In recent years, with the development of novel sequencing technologies and bioinformatics, genome-wide WRKY

analysis has been performed in many plant species including *Populus trichocarpa* (He et al., 2012), *Pyrus*

bretschneideri (Huang et al., 2015), *Citrus* (Ayadi et al., 2016), *Glycine max* (Song et al., 2016), *Daucus carota* (Li

et al., 2016) and so on. Studies on WRKY identification and functional analysis have shown that WRKY TFs play

76 significant roles in signaling and regulation of expression during various biotic and abiotic stresses. In banana,

77 MaWRKY26 activated jasmonic acid biosynthesis and enhanced cold tolerance in the fruit (Ye et al., 2016). In

78 wheat, *TaWRKY49* and *TaWRKY62* were shown to participate in the defense response against the fungal pathogen

79 *Puccinia striiformis* f. sp tritici (Pst), *TaWRKY49* was shown to be a negative regulator and *TaWRKY62* a positive

80 regulator of wheat's HTSP resistance to Pst (Wang et al., 2017). WRKY TFs have also been implicated in the

81 modulation of plant development. In the poplar *Populus trichocarpa*, *PtrWRKY19* may function as a negative

82 regulator of pith secondary wall formation (Yang et al., 2016). In foxtail millet, map-based cloning, combined with

83 high-throughput sequencing, revealed that LP1, which encodes a novel WRKY TF, regulates panicle development

84 (Xiang et al., 2017). WRKY TFs have also been shown to regulate the production of several secondary metabolites

85 such as phenolic compounds including lignin, flavanols and tannins. In *Arabidopsis*, *AtWRKY23* regulates the

86 production of flavanols in auxin inducible manner (Grunewald et al., 2008; 2013). In rice, *OsWRKY76* activates cold

87 stress tolerance but suppresses PR genes and production of phytoalexins like terpene and the phenylpropanoid

88 sakuranetin (Yokotani et al., 2013). In *Withania somnifera*, WsWRKY1 binds to W-box sequences in promoters

89 encoding squalene synthase and squalene epoxidase, indicating that it has a direct role in the regulation of the

90 triterpenoid pathway (Singh et al., 2017). What's more, the WRKYs always work interaction with other proteins,

91 such as *PeWRKY83* could interact with *PeVQ* proteins in moso bamboo (Wu et al., 2017) and physical interaction of

92 WRKY75 with DELLA repressors were also found in *Arabidopsis thaliana* (Zhang et al., 2018).

93 *Moringa oleifera* Lam., commonly known as drumstick, belongs to the monogeneric family *Moringaceae*

94 (Ramachandran, 1980). This species is widely cultivated in tropical and sub-tropical areas and has a long history of

95 traditional medicine and culinary uses (Anwar et al., 2007; Zhang et al., 2017). Drumstick is considered to be a fast-

96 growing tree species and also it's a drought tolerant plant that can be grown in diverse soils except those that are

97   waterlogged; it may also become important for biofuel production and has been used in a variety of industrial

98   applications (Popoola and Obembe, 2013; Shih et al., 2011). Studies on drumstick transcription factors have hitherto

99   rarely been reported because of a lack of genomic data for this species. The publication of the drumstick genome

100   draft database (Tian et al., 2015) provides resources with which to carry out bioinformatics-based identification and

101   analysis of WRKY TFs. In the present study, we have used these genomic resources to identify members of the

102   WRKY gene family in drumstick and correlated their expression with various stress responses. We carried out a

103   detailed study of the drumstick WRKY gene family, including gene classification, phylogenetic analysis,

104   determination of structural organization and conserved motif composition, and assessed the selective pressures that

105   have acted on different members of this family.

106   **Material and methods**

107   **Sequence database searches**

108   The complete genome and proteome sequences and General Feature Format (GFF) file for *Arabidopsis* were

109   downloaded from TAIR (http://www.arabidopsis.org). The annotated drumstick genome sequences were provided

110   by Yunnan Agricultural University. A WRKY-domain Hidden Markov Model (HMM) Profile, which was

111   downloaded from Pfam (http://pfam.xfam.org/), was used as a query with which to search all of the annotated

112   proteins in the drumstick genome with an E value cut-off of 1E-5. The candidates selected using HMMER were

113   examined to determine whether they had typical features of WRKY proteins by employing the Pfam database.

114   Finally, the CD-HIT program and the Pfam database were used to eliminate duplicate and incomplete sequences.

115   Non-overlapping WRKY protein sequences were used for further analysis.

116   **Multiple sequence alignment and phylogenetic analyses**

117   The conserved WRKY domains of *MoWRKY* genes obtained using manual inspection in the Pfam program were

118   aligned using ClustalX 1.83 software. Phylogenetic analysis including 7 representative domains from *Arabidopsis*

119   was carried out to obtain better classifications of the different clades by applying the Neighbor-Joining method with

120   1000 bootstrap replicates using MEGA 6 software.

121   **Gene structure and motif composition analysis**

122   Analysis of the exon-intron organization of *Mo*WRKYs was performed by comparing the coding sequences of

123   *Mo*WRKYs with their corresponding genomic sequences using GSDS software (http://gsds.cbi.pku.edu.cn).

124   Conserved motifs in each WRKY protein were investigated using the Multiple Expectation Maximization for Motif

125   Elucidation (MEME) online program: http://meme-suite.org/. The following parameters were employed in analysis:

126   maximum number of motifs 20; minimum motif width 6; maximum motif width 50.

127   **Promoter *cis-acting* elements analysis of *MoWRKYs***

128   The promoter sequences, 1.5kb upstream of the translation start site, of the *MoWRKY* genes were obtained from

129   drumstick genome. PlantCARE (Lescot et al., 2002) was used to analyse the *MoWRKY* gene promoters and identify

130   their cis-acting elements.

131   **Tests for selective pressure**

132   The multiple sequence alignment of drumstick MoWRKY proteins was carried out using ClustalW with default

133   parameters. Then the sequences were trimmed to reduce gap penality. DNAMAN was used to search for nucleotide

134   sequences encoding additional WRKY proteins, with the aligned MoWRKY protein sequences as guides. The

135   synonymous (K_S) and nonsynonymous (K_a) substitution rates were calculated with the YN00 program in PAML4.9

136   with default parameters (Yang et al., 2007).

137   **Expression analysis**

138   To investigate the patterns of expression of *MoWRKY* genes under normal and abiotic stress conditions, seedlings

139   of drumstick were cultivated in potting soil at 25°C under 14: 10h light: dark conditions in a growth chamber for 20

140   days before treatment. For salt and oxidative stress treatments, seedlings were sprayed for 12h with, respectively,

141   150mM NaCl and H_2O_2 solution. Cold and heat stress were applied by transferring plants to a climate chamber at,

142   respectively, 4°C and 42°C for 12h. Drought stress was induced by withholding water for 2 weeks. Each treatment

143   consisted of three replicates. After stress treatments, total RNA was isolated from leaf, stem, stem tip and root

144   tissues of each seedling using a Total RNA Kit (OMEGA, Guangzhou, China). Total RNA was reverse transcribed

145   into cDNA using a PrimeScript RT Master Mix (Perfect) Real Time Kit (Takara, Dalian, China). Gene specific

146   primers were designed using Primer 5.0 and the *RPL* gene was used as a reference (Deng et al., 2016). Expression of

147   all *MoWRKY* genes was examined by RT-PCR and products from each sample were analyzed using a 1% agarose

148   gel. Among all *MoWRKY* genes, 9 genes belonging to different subgroups were selected for analysis of gene

149   expression levels using qRT-PCR according to the method described in Wei et al. (2016) and *RPL* was amplified as

150   a reference gene (Deng et al., 2016). Relative expression levels were evaluated using the $2^{-\Delta\Delta CT}$. Three technical

151   replicates were conducted for test and reference genes of each sample to obtain precise and reproducible results.

152   Statistical analysis was carried out using SPSS 19.0 software (SPSS Inc., Chicago, IL, USA), Duncan's multiple

153   range test was used to detect differences among means. A *p*-value < 0.05 was considered significant.

154   **Results**

155   **Identification of WRKY family members in the drumstick genome**

156   To identify all the *WRKY* genes in the drumstick genome, we employed the HMM profile of the WRKY domain

157   (PF03106) as a query to search against the drumstick genome database using HMMER 3.0 and BLAST. A total of

158   54 nonredundant genes (Table 1) were identified as *WRKY* genes and a unique name was assigned to each drumstick

159   *WRKY* gene, consisting of two italic letters denoting the source organism and sequential numbers: *MoWRKY1* to

160   *MoWRKY54*. All the putative 54 *WRKY* genes were further analyzed to confirm the presence of the WRKY domain

161   and all of them were annotated with gene ontology (GO) terms (Supplementary file 1). Fifty-three *MoWRKY* genes

162   containing complete WRKY domains were identified; only one gene (*MoWRKY50*) lacked a complete domain. The

163   highly conserved domain WRKYGQK was present in 52 of the MoWRKY proteins, whereas the remaining one

164   (MoWRKY24) contained a WRKYGKK domain. The lengths of the MoWRKY proteins ranged from 106

165   (*MoWRKY24*) to 834 (*MoWRKY3*) amino acids; the average length was 391 amino acids.

166   **Phylogenetic relationship and classification of *MoWRKY* genes**

167   The most prominent structural feature of WRKY genes is a conserved WRKY domain; there is also a zinc-finger

168   motif. Among the 54 MoWRKY proteins identified, nine MoWRKY proteins contained two WRKY domains; since

169   one MoWRKY protein did not have a complete WRKY domain, a total of 62 WRKY domains were found in this

170   study. In each protein that contained two WRKY domains, we designated these domains by the WRKY name plus N

171   or C for the N-terminal or C-terminal domain respectively. In order to examine phylogenetic relationships and

172   classify all 62 MoWRKY domains, a phylogenetic tree based on conserved WRKY domains was constructed.

173   Representative WRKY domains from Arabidopsis were used in our analysis, and the candidate domains were

174   obtained from Diao (2016) and Li (2016). Fig. 1 shows a multiple sequence alignment of the 62 WRKY domains.

175   Three major groups were identified, as previously described in poplar (He et al., 2012), pepper (Diao et al., 2016)

176   and carrot (Li et al., 2016). Additionally, several subgroups were apparent on the basis of the phylogenetic analysis.

177   Group I contained 10 WRKY proteins, of which all contain two WRKY domains except for MoWRKY10. This

178   member might have lost the N-terminal WRKY domain during evolution, since its single WRKY domain showed

179   high similarity to MoWRKY1C, which is located in the C-terminal WRKY domain clade, suggesting a common

180   origin for these two domains. Group II had the largest numbers of WRKY proteins and was divided into five major

181   subgroups: IIa, IIb, IIc, IId and IIe. Subgroup IIa (3 members) and IIb (8 members) were two subgroups in the same

182   branch, while subgroup IId (5 members) and IIe (7 members) were derived from one clade. Subgroup IIc, with 14

183   members, was more similar to group I than to any other subgroups according to the phylogenetic analysis.

184   Furthermore, 6 WRKY domains belonged to group III, which is widely considered to be the most advanced in terms

185   of evolution and the most relevant to adaptability (Dou et al., 2016; Kalde et al., 2003; Huang et al., 2016).

186   Comparing the two phylogenetic trees, constructed for MoWRKY domains and genes, similar groups and subgroups

187   were identified, though the classifications of a few members were different (Fig. 1 and Fig. S1), indicating that the

188   conserved WRKY domain is an important unit in WRKY proteins.

189   **Structure analysis of *MoWRKY* genes**

190   Intron/exon organization and numbers of introns are typical imprints of evolution within some gene families. In

191   this study, we analysed the structure of *MoWRKY* genes to gain further insight into evolutionary events that had

192   shaped them and found that all *MoWRKY* genes contain introns (Fig. 2 A). The number of introns varies among

193   genes, with the minimum, one intron, identified in five *MoWRKY*s (*MoWRKY50*, *MoWRKY44*, *MoWRKY51*,

194   *MoWRKY19* and *MoWRKY11*) of subgroup IIc and the maximum, 10 introns, being present in *MoWRKY22*. Gene

195   structure analysis revealed that genes with similar structures always clustered in the same class. For example, six

196   members of group III all contained 3 exons and 2 introns. Similarly, 5 exons and 4 introns were present in

197   *MoWRKY2*, *MoWRKY3*, *MoWRKY4*, *MoWRKY5* and *MoWRKY6*, which belonged to group I. However, the other 5

198   *MoWRKY*s in group I exhibited different gene structures.

199   **Motif composition analysis of MoWRKY proteins**

200   The conserved motifs of WRKY proteins in drumstick were investigated using the MEME online software suite

201   (http://meme-suite.org/) to better understand the similarity and diversity of motif compositions. Twenty distinct

202   motifs were identified and a schematic overview of these motifs is provided in Fig. 3. For MoWRKY proteins, motif

203   1 was broadly distributed in all MoWRKY proteins, which was corresponded to WRKY domain. Motif 3 was only

204   detected in the type I group. Motifs 5, 6, 7, 8, 9, 16, 17, 18 and 20 were only detected in the type II group; among

205   them, motifs 5 and 6 were only detected in subgroup IIa and IIb, motifs 8 and 9 were only detected in subgroup IIb,

206   motif 17 was only detected in subgroup IIc, and motif 16 was only detected in subgroup IId. Motifs 12 and 15 were

207   only detected in the type III group. Generally, proteins with similar motif compositions were clustered in the same

208   class indicating that members of the same class may have similar functions.

209   **Rapid expansion of group III WRKY genes in land plants**

210   Group III WRKY genes have only been characterized in flowering plants, and a large number of duplications and

211   diversifications in this group appear to have resulted from different selection challenges (Dou et al., 2016; Huang et

212   al., 2016; Kalde et al., 2003). To explore the evolutionary relationships of group III WRKY genes across drumstick

213   and other land plant species, we performed a multiple sequence alignment among the 81 group III WRKY proteins

214   from drumstick and another 7 species. A phylogenetic tree was constructed from the results of the alignment using

215   the neighbor-joining method (Fig. 4). The marked difference in group III WRKY gene size among different species

216   suggests that group III WRKY gene expansion occurred after the divergence of monocotyledons and dicotyledons.

217   MoWRKY clearly shared more sequence similarity with VvWRKY and PaWRKY than with other WRKYs.

218   **The *cis-acting* elements analysis of *MoWRKYs***

219   For further understand the possible functions of *MoWRKY* genes, the cis-acting elements in all *MoWRKY* genes

220   promoters were analyzed using PlantCARE software based the drumstick genome data. Various types of *cis-acting*

221   elements were found and all *MoWRKY* genes contained several *cis-acting* elements in their promoter regions. The 10

222   most common elements were summarized in Table 2. These elements included three hormone responsive elements

223   (ABRE, CGTCA motif and TGACG motif), an essential element for the anaerobic inductio (ARE), a drought stress

224   responsive element (MBS), a heat stress responsive element (HSE) and four light responsive elements (Sp1, Box 4,

225   G box and GT1 motif).

226   **Divergence in selective pressure between subgroups**

227   The ratio (ω) of the non-synonymous substitution rate (Ka) to the synonymous substitution rate (K_S) provides a

228   sensitive measure of selective pressure acting on a protein-coding gene. Homologous genes with ω ratios of 1, <1, or

229   >1 are usually assumed to be evolving under neutral evolution, purifying selection, or positive selection,

230   respectively. To test for deviations in the substitution rates of *MoWRKY* genes, we calculated ω values across all

231   pairwise comparisons within the 54 *WRKY* genes using the YN00 program in the PAML software package. The

232   frequency distribution of ω values is shown in Fig. 5 A. The results suggested that the *WRKY* gene family evolved

233   mainly under strong purifying selection. However, there are several ω values greater than 1, such as those for the

234   comparison between *MoWRKY8* and *MoWRKY9* and that between *MoWRKY8* and *MoWRKY10*, indicating that

235   positive selection acted on these genes. Only 0.5% of the ω values approximated to 1, indicating that no selective

236   pressure acted on these genes.

237   To test whether the rate of evolution among the subgroups of *WRKY* genes was identical, we calculated ω values

238   across all pairwise comparisons within each of the subgroups; the results are shown in Fig. 5 B. The average ω

239   values of each subgroup were different. In order (highest first) they were: IIc, III, I, IIe, IId, IIb and IIa, indicating

240   that different subgroups had evolved at different rates and that IIc had evolved the fastest.

241   **Expression patterns of WRKY genes in drumstick under normal growth conditions and abiotic stress**

242   **conditions**

243   To investigate the responses of *MoWRKY* genes to stresses, we examined the expression patterns of all 54 full-

244   length *MoWRKY*s under normal growth conditions and under five abiotic stresses (heat, cold, drought, salt and

245   oxidative) in different tissues (leaves, roots, stems, stem apex) using RT-PCR. As shown in Fig. 2 B, among the 54

246   *MoWRKY* genes, 13 genes were expressed in all tissues under all growth conditions. In contrast, six genes, including

247   *MoWRKY24*, the only gene with a variant WRKY domain (WRKYGKK), were not expressed in any tissue or in

248   response to any of the treatments applied in this study. Thus, these 6 *WRKY* genes are expressed at undetectably low

249   levels, or they are only induced in response to treatments and/or in tissues not examined in our study, or they are

250   pseudogenes. The other 35 *WRKY* genes were expressed selectively in a specific tissue and/or in response to a

251   specific treatment. Six of these genes were not expressed in any tissue under normal growth conditions but were

252   expressed under stress conditions, suggesting that they play specific roles during stress conditions. At the same time,

253   some genes, such as *MoWRKY46*, were only expressed in specific tissues under normal growth conditions but were

254   expressed in all tissues under certain stress conditions, indicating that these genes may also play specific roles under

255   stress conditions.

256   Nine *MoWRKY* genes from different subgroups were selected and their expression profiles were analyzed in root

257   tissue under normal growth conditions and five abiotic stresses using qRT-PCR. As shown in Fig. 6, these selected

258   *MoWRKY* genes were sensitive to abiotic stresses. All 9 exhibited a high level of transcript accumulation under cold

259   stress, especially *MoWRKY30* (GO: 0006950), followed by *MoWRKY54* (GO: 0006950 and GO: 0080134).

260   Interestingly, the genes that were most strongly up-regulated under cold treatment were always up-regulated in

261   response to heat and salt treatments. In drought stress, *MoWRKY22*, which had the most introns and *MoWRKY3*,

262   which was the longest in *MoWRKY* gene family were found to be slightly upregulated, whereas weak expression

263   were found for the other seven genes. The expression levels of almost all the nine *MoWRKY*s were decreased under

264   oxidative stress. *MoWRKY49*, *MoWRKY53* and *MoWRKY54*, which all belonged to group III, have similar gene

265   structures and the same motifs. But the expression levels of the three genes under abiotic stresses were slightly

266   different. They were evidently upregulated in cold and salt to different degrees; *MoWRKY53* (GO: 0006950 and GO:

267   0080134) and *MoWRKY54* (GO: 0006950 and GO: 0080134) were also responsive to heat. Overall, the expression

268   patterns of *MoWRKY*s under various conditions suggest that different *MoWRKY* genes may be involved in different

269   signaling and stress responses, and that an individual *MoWRKY* gene can also participate in multiple signaling and

270   stress process.

271   **Discussion**

272   WRKY transcription factors were first identified over 20 years ago (Ishiguro and Nakamura, 1994) and it has

273   been suggested that they play important roles in stress responses and at many stages of plant growth and

274   development (Phukan et al., 2016; Tripathi et al., 2014). Genes encoding WRKY proteins belong to a large family,

275   with 72 members in *Arabidopsis thaliana* (Wang et al., 2011), 100 in *Oryza sativa* (Ross et al., 2007) and 104

276   members in the *Populus trichocarpa* genome (He et al., 2012). A previous study showed that *Populus trichocarpa*

277   (He et al., 2012) and *Daucus carota* (Li et al., 2016) WRKYs could be divided into three groups. In the present

278   study, when a phylogenetic tree of WRKYs from drumstick and *Arabidopsis* was constructed, we found that the 54

279   WRKYs from drumstick fell into three distinct groups. This result was consistent with the WRKY domain and zinc

280   finger type classification of these WRKYs. When the subgroups of WRKY genes were compared among

281   *Arabidopsis*, rice and poplar, we found that the number of each subgroup in group II was similar indicating that all

282   members of these subgroups have probably been identified. However, the number of *MoWRKY*s in group III is less

283   than the numbers in *Arabidopsis* and rice which are older species, implying that *WRKY* genes of this group in

284   drumstick either had been lost during the course of evolution or were underrepresented in our analysis.

285   The WRKY conserved domain is the most important functional and evolutionary unit of WRKY transcription

286   factors. Although the WRKYGQK amino acid residues in the WRKY domain are highly conserved, there are

287   variants. Six sequence variations (WRKYGHK, WRKYGQN, WRKYGKK, WRKCGQK, WRKYGQT,

288   WRKYGMK) were found in *CaWRKY* genes. Six heptapeptide variants, namely WRKYGKK, WRKYGEK,

289   WRKYGKR, WRKYEDK, WKKYGQK, WHQYGLK, were found in soybean (Song et al., 2016). In our study,

290   only one variant (WRKYGKK) was found, and only in *MoWRKY24* which belongs to subgroup IIc. WRKYGKK is

291   the most common variant in many species. In the tobacco WRKY protein family, the WRKYGKK domain could

292   bind specifically to a WK-box, which was significantly different from the W-box (Verk et al., 2008). In our study,

293   we could not detect the expression of *MoWRKY24* in any tissues or under any stress conditions. The reason may be

294   that the expression level of *MoWRKY24* was too low to be detected, or that this gene is only expressed under special

295   conditions, or that it has become a pseudogene. This apparent lack of expression needs to be investigated further.

296   The structures of the *MoWRKY* genes showed group-specific exon-intron patterns, as is also the case in carrot (Li

297   et al., 2016) and cassava (Wei et al., 2016). Exon-intron structural diversity plays an important part in the evolution

298   of gene families (Wei et al., 2016). The number of introns in *MoWRKY* genes varied from 1 to 10. However, in

299   poplar (He et al., 2012) and cassava (Wei et al., 2016), the number of introns varied from**,** respectively, 0 to 6 and 1

300   to 5. The results indicated that *MoWRKY*s have more gene structure diversity than the poplar and cassava *WRKY*

301   genes. In our study, the length of the *MoWRKY3* gene in group I was greater than those of any other genes. While

302   neither the number nor the length of exons in this gene was unusually high, there were more introns. Combined

303   motif compositions, we can find the variety and average length of motifs identified in MoWRKY3 were not

304   especially large, indicating that their functions were probably not influenced by the presence of the numerous introns.

305   According to a previous report, the rate of intron loss is faster than the rate of intron gain after segmental duplication

306   (Nuruzzaman et al., 2010) and intron loss can result from intron turnover or reverse transcription of the mature

307   mRNA followed by homologous recombination with intron-containing alleles (He et al., 2012). In drumstick,

308   members of group III all contained two introns; the average number of introns in the other groups was more than

309   that in this group. Consequently, it can be inferred that group III developed later than other groups. The structure

310   and motif compositions of group III members were very similar, indicating that these genes expanded not by

311   merging, transfer or loss but in other ways.

312   WRKY proteins usually functioned as transcriptional regulators by binding to W-box to regulate defense-related

313   genes. In our study, we found that nearly half *MoWRKY* genes also contained W-box element in their promoter

314   regions. The same findings were identified in carrot (Li et al., 2016) and soybean (Song et al., 2016), suggesting that

315   these *MoWRKY* genes are auto-regulated by themselves or cross-regulated. Accumulating evidence suggests that

316   WRKY transcription factors are involved in many plant processes including development and responses to biotic

317   and abiotic stresses and that may due to the upstream genes specificity bind the corresponding cis element to

318   regulate the expression of *WRKY* genes. In carrot, fourteen selected *DcWRKY* genes responded to whitefly and aphid

319   infections and twelve *DcWRKY* genes were upregulated or downregulated under heat and/or cold treatments (Li et

320   al., 2016). At least 31 *PeWRKY* genes in moso bamboo (Li et al., 2017) and 21 *CaWRKY* genes in pepper (Diao et

321   al., 2016) were differentially expressed under abiotic stresses. Similarly, 55 *VvWRKY* genes in grape (Zhang and

322   Feng, 2014) differentially responded to at least one abiotic stress treatment. In our study, the results of expression

323   pattern analysis demonstrated that most *MoWRKY* genes had different expression levels when the seedlings were

324   exposed to different stresses despite highly homologous amino acid sequences and conserved domain structures.

325   *WRKY* genes within the same group may act as redundant and substitute members in regulating functions. The very

326   large expression differences suggested that the products of these genes have different physiological functions,

327   facilitating adaptation to complex challenges. Further structural analyses and investigations into the expression

328   patterns of the *MoWRKY* gene family would facilitate a more comprehensive understanding of the specific functions

329   of individual WRKY genes. The current investigation highlights a number of *MoWRKY* genes that may be involved

330   in stress defenses, and lays a solid foundation for the selection of candidate genes for further studies.

331   **Conclusion**

332   The publication of drumstick genome sequences provides an opportunity for genome wide identification and

333   characterization of WRKY TFs. Bioinformatics tools have been made in the present study to identify the putative

334   members of WRKY genes of drumstick and subject it to characterization for gene structures, motif analysis,

335   conserved motifs and phylogenetic tree construction. The multiple members of *WRKY* genes in plants reflect the

336   redundancy and differentiated functions of these proteins which need to be explored by expression profiling. The

337   expression profiling under different abiotic stress conditions revealed several potential MoWRKYs showing higher

338   expression level under drought, salt, cold and heat stresses.

339   **Acknowledgements**

340   Authors would like to thank professor Jun Sheng and Yang Tian in Yunnan Agricultural University for their kind

341   sharing drumstick genome sequences data. And authors also thank to the reviewers for their careful reviewing and

342   helpful comments on the manuscript.

343

344   **References**

345   Anwar F, Latif S, Ashraf M, Gilani AH. 2007. *Moringa oleifera*: A food plant with multiple medicinal uses.

346   Phytother Res 21:17-25.

347   Ayadi M, Hanana M, Kharrat N, Merchaoui H, Ben Marzoug R, Lauvergeat V, Rebai A, Mzid R. 2016. The WRKY

348   transcription factor family in Citrus: valuable and useful candidate genes for Citrus breeding. Appl Biochem Biotech

349   180:516-543.

350   Chen LG, Song Y, Li SJ, Zhang LP, Zou CS, Yu DQ. 2012. The role of WRKY transcription factors in plant abiotic

351   stresses. Biochim Biophys Acta 1819:120-128.

352   Deng LT, Wu YL, Li JC, Ouyang KX, Ding MM, Zhang JJ, Li SQ, Lin MF, Chen HB, Hu XS. 2016. Screening

353   reliable reference genes for RT-qPCR analysis of gene expression in *Moringa oleifera*. Plos One. DOI:

354   10.1371/journal.pone.0159458.

355   Diao WP, Snyder JC, Wang SB, Liu JB, Pan BG, Guo GJ. 2016. Genome-Wide identification and expression

356   analysis of WRKY gene family in Capsicum annuum L.. Front Plant Sci. DOI: 10.3389/fpls.2016.01727.

357   Dou LL, Guo YN, Evans O, Pang CY, Wei HL, Song MZ, Fan SL, Yu SX. 2016. Identification and expression

358   analysis of group III WRKY transcription factors in cotton. J Integr Agric 15:2469-2480.

359   Eulgem T, Rushton PJ, Robatzek S, Somssich IE. 2000. The WRKY superfamily of plant transcription factors.

360   Trends Plant Sci 5:199-206.

361   Goel R, Pandey A, Trivedi PK, Asif MH. 2016. Genome-Wide analysis of the Musa WRKY gene family: evolution

362   and differential expression during development and stress. Front Plant Sci. DOI: 10.3359/fpls.2016.00299.

363   Grunewald W, De Smet I, De Rybel B, Robert HS, Van De Cotte B, Willemsen V, Gheysen G, Weijers D, Friml J,

364   Beeckman T. 2013. Tightly controlled WRKY23 expression mediates Arabidopsis embryo development. Embo Rep

365   14:1136-1142.

366   Grunewald W, Karimi M, Wieczorek K, Van de Cappelle E, Wischnitzki E, Grundler F, Inze D, Beeckman T,

367   Gheysen G. 2008. A role for AtWRKY23 in feeding site establishment of plant-parasitic nematodes. Plant Physiol

368   148:358-368.

369   He HS, Dong Q, Shao YH, Jiang HY, Zhu SW, Cheng BJ, Xiang Y. 2012. Genome-wide survey and

370   characterization of the WRKY gene family in *Populus trichocarpa*. Plant Cell Rep 31:1199-1217.

371   Huang XS, Li KQ, Xu XY, Yao ZH, Jin C, Zhang SL. 2015. Genome-wide analysis of WRKY transcription factors

372   in white pear (*Pyrus bretschneideri*) reveals evolution and patterns under drought stress. BMC Genomics. DOI:

373   10.1186/s12864-015-2233-6.

374   Huang Y, Li MY, Wu P, Xu ZS, Que F, Wang F, Xiong AS. 2016. Members of WRKY Group III transcription

375   factors are important in TYLCV defense signaling pathway in tomato (*Solanum lycopersicum*). BMC Genomics.

376   DOI: 10.1186/s12864-016-3123-2.

377   Kalde M, Barth M, Somssich IE, Lippok B. 2003. Members of the Arabidopsis WRKY group III transcription

378   factors are part of different plant defense signaling pathways. Mol Plant-Microbe Interact 16:295-305.

379   Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. 2002. PlantCARE: a

380   database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences.

381   Nucleic Acids Res 30:325-327.

382   Li L, Mu SH, Cheng ZC, Cheng YW, Zhang Y, Miao Y, Hou CL, Li XP, Gao J. 2017. Characterization and

383   expression analysis of the WRKY gene family in moso bamboo. Sci Rep. DOI: 10.1038/s41598-017-06701-2.

384   Li MY, Xu ZS, Tian C, Huang Y, Wang F, Xiong AS. 2016. Genomic identification of WRKY transcription factors

385   in carrot (*Daucus carota*) and analysis of evolution and homologous groups for plants. Sci Rep. DOI:

386   10.1038/srep23101.

387   Nuruzzaman M, Manimekalai R, Sharoni AM, Satoh K, Kondoh H, Ooka H, Kikuchi S. 2010. Genome-wide

388   analysis of NAC transcription factor family in rice. Gene 465:30-44.

389   Pandey SP, Somssich IE. 2009. The role of WRKY transcription factors in plant immunity. Plant physiol 150:1648-

390   1655.

391   Phukan UJ, Jeena GS, Shukla RK. 2016. WRKY transcription factors: molecular regulation and stress responses in

392   plants. Front Plant Sci. DOI: 10.3389/fpls.2016.00760.

393   Popoola JO, Obembe OO. 2013. Local knowledge, use pattern and geographical distribution of *Moringa oleifera*

394   Lam. (*Moringaceae*) in Nigeria. J Ethnopharmacol 150:682-691.

395   Ramachandran C, Peter KV, Gopalakrishnan PK. 1980. Drumstick (*Moringa oleifera*): A multipurpose Indian

396   vegetable. Econ Bot 34:276-283.

397   Ross CA, Liu Y, Shen QJ. 2007. The WRKY gene family in Rice (*Oryza sativa*). J Integr Plant Biol 49:827-842.

398   Shih MC, Chang CM, Kang SM, Tsai ML. 2011. Effect of different parts (leaf, stem and stalk) and seasons (summer

399   and winter) on the chemical compositions and antioxidant activity of *Moringa oleifera*. Int J Mol Sci 12:6077-6088.

400   Singh AK, Kumar SR, Dwivedi V, Rai A, Pal S, Shasany AK, Nagegowda DA. 2017. A WRKY transcription factor

401   from *Withania somnifera* regulates triterpenoid withanolide accumulation and biotic stress tolerance through

402   modulation of phytosterol and defense pathways. New Phytol 215:1115-1131.

403   Smith NC, Matthews JM. 2016. Mechanisms of DNA-binding specificity and functional gene regulation by

404   transcription factors. Curr Opin Struct Biol 38:68-74.

405   Song H, Wang PF, Hou L, Zhao SZ, Zhao CZ, Xia H, Li PC, Zhang Y, Bian XT, Wang XJ. 2016. Global analysis of

406   WRKY genes and their response to dehydration and salt stress in Soybean. Front Plant Sci. DOI:

407   10.3389/fpls.2016.00009.

408   Tian Y, Zeng Y, Zhang J, Yang CG, Yan L, Wang XJ, Shi CY, Xie J, Dai TY, Peng L, Zeng HY, Xu AN, Huang

409   YW, Zhang JJ, Ma X, Dong Y, Hao SM, Sheng J. 2015. High quality reference genome of drumstick tree (*Moringa*

410   *oleifera* Lam.), a potential perennial crop. Sci China 58:627-638.

411   Tripathi P, Rabara RC, Rushton PJ. 2014. A systems biology perspective on the role of WRKY transcription factors

412   in drought responses in plants. Planta 239:255-266.

413   Ulker B, Somssich IE. 2004. WRKY transcription factors: from DNA binding towards biological function. Curr

414   Opin Plant Biol 7:491-498.

415   Verk MCV, Pappaioannou D, Neeleman L, Bol JF, Linthorst HJM. 2008. A novel WRKY transcription factor is

416   required for induction of PR-1a gene expression by salicylic acid and bacterial elicitors. Plant Physiol 146:1983-

417   1995.

418   Wang JJ, Tao F, Tian W, Guo ZF, Chen XM, Xu XM, Shang HS, Hu XP. 2017. The wheat WRKY transcription

419   factors TaWRKY49 and TaWRKY62 confer differential high-temperature seedling-plant resistance to *Puccinia*

420   *striiformis* f. sp. *tritici*. Plos One. DOI: 10.1371/journal.pone.0181963.

421   Wang QS, Wang MH, Zhang XZ, Hao BJ, Kaushik SK, Pan YC. 2011. WRKY gene family evolution in

422   *Arabidopsis thaliana*. Genet 139:973-983.

423   Wei YX, Shi HT, Xia ZQ, Tie WW, Ding ZH, Yan Y, Wang WQ, Hu W, Li KM. 2016. Genome-wide identification

424   and expression analysis of the WRKY gene family in Cassava. Fron Plant Sci. DOI: 10.3389/fpls.2016.00025.

425   Wu M, Liu HL, Han GM, Cai RH, Pan F, Xiang Y. 2017. A moso bamboo WRKY gene PeWRKY83 confers

426   salinity tolerance in transgenic Arabidopsis plants. Sci Rep. DOI: 10.1038/s41598-07-10795-z.

427   Xiang JS, Tang S, Zhi H, Jia GQ, Wang HJ, Diao XM. 2017. Loose panicle1 encoding a novel WRKY transcription

428   factor, regulates panicle development, stem elongation, and seed size in foxtail millet [*Setaria italica* (L.) P. Beauv.].

429   Plos One. DOI: 10.1371/journal.pone.0178730.

430   Yang L, Zhao X, Yang F, Fan D, Jiang YZ, Luo KM. 2016. PtrWRKY19, a novel WRKY transcription factor,

431   contributes to the regulation of pith secondary wall formation in Populus trichocarpa. Sci Rep. DOI:

432   10.1038/srep18643.

433   Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591.

434   Ye YJ, Xiao YY, Han YC, Shan W, Fan ZQ, Xu QG, Kuang JF, Lu WJ, Lakshmanan P, Chen JY. 2016. Banana

435   fruit VQ motif-containing protein5 represses cold-responsive transcription factor MaWRKY26 involved in the

436   regulation of JA biosynthetic genes. Sci Rep. DOI: 10.1038/srep23632.

437   Yokotani N, Sato Y, Tanabe S, Chujo T, Shimizu T, Okada K, Yamane H, Shimono M, Sugano S, Takatsuji H,

438   Kaku H, Minami E, Nishizawa Y. 2013. WRKY76 is a rice transcriptional repressor playing opposite roles in blast

439   disease resistance and cold stress tolerance. J Exp Bot 64:5085-5097.

440   Zhang LP, Chen LG, Yu DQ. 2018. Transcription Factor WRKY75 Interacts with DELLA Proteins to Affect

441   Flowering. Plant Physiol 176, 790-803.

442   Zhang JJ, Yang YS, Lin MF, Li SQ, Tang Y, Chen HB, Chen XY. 2017. An efficient micropropagation protocol for

443   direct organogenesis from leaf explants of an economically valuable plant, drumstick (*Moringa oleifera* Lam.). Ind

444   Crops Prod 103:59-63.

445   Zhang Y, Feng JC. 2014. Identification and Characterization of the Grape WRKY Family. Biomed Res Int. DOI:

446   10.1155/2014/787680.

447   Zhang YJ, Wang LJ. 2005. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in

448   plants. BMC Evol Biol. DOI: 10.1186/1471-2148-5-1.

449

450

451

452   **Caption**

453   **Fig. 1** Phylogenetic tree of the WRKY conserved domain from drumstick and selected *Arabidopsis*. The

454   bootstrap test was performed with 1000 replicates.

455   **Fig. 2** Exon-intron composition and expression patterns of *MoWRKY* genes. (A) Exon-intron composition. (B)

456   Expression patterns.

457   **Fig. 3** Distribution of conserved motifs in MoWRKYs. Different colors represent different motifs.

458   **Fig. 4** Phylogenetic tree of 81 group III WRKY proteins from drumstick and other seven species. *O. sativa*

459   (blue triangles and lines), *P. euphratica* (green lines), *V. vinifera* (orange lines), *P. patens* (purple triangle), *A.*

460   *thaliana* (brown triangle and lines), *S. moellendorfii* (grass green lines) and *P. abies* (red lines).

461   **Fig. 5** Frequency distribution and average values of Ka/Ks ratios. (A) Frequency distribution between any two

462   drumstick WRKY genes. (B) Average values of Ka/Ks across sub-groups of drumstick WRKYs.

463   **Fig. 6** Expression profiles for 9 selected *MoWRKY* genes in root under different stresses. (A) *I-MoWRKY3*; (B)

464   *IIa-MoWRKY22*; (C) *IIb-MoWRKY30*; (D) *IIe-MoWRKY50*; (E) *IId-MoWRKY27*; (F) *IIe-MoWRKY35*; (G) *III-*

465   *MoWRKY49*; (H) *III-MoWKRY53*; (I) *III-MoWKRY54*.

466   **Fig. S1** Phylogenetic tree of MoWRKYs.

467

Figure 1

Phylogenetic tree of the WRKY conserved domain from drumstick and selected Arabidopsis.

The bootstrap test was performed with 1000 replicates.

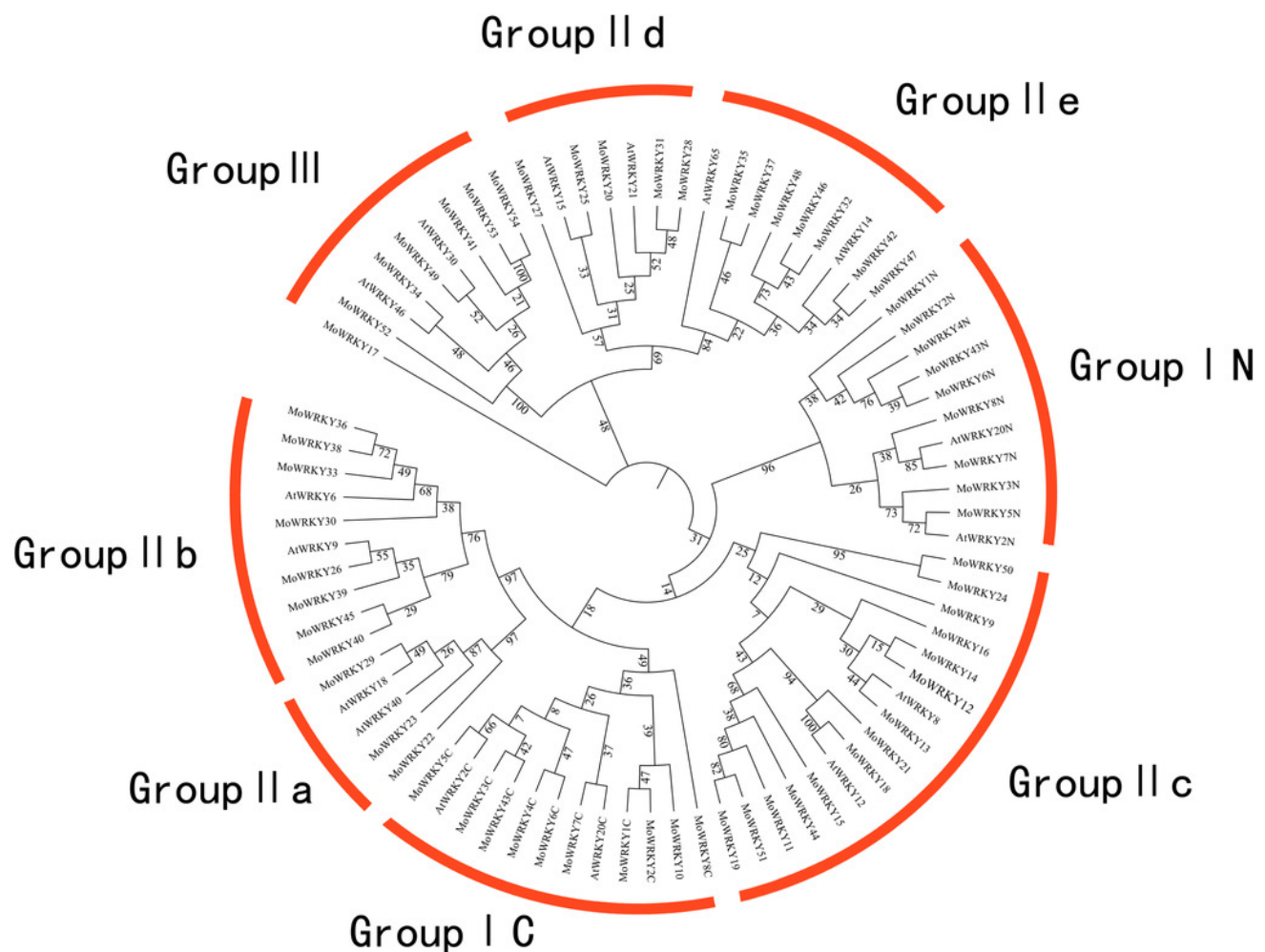



Figure 2

Exon-intron composition and expression patterns of *MoWRKY* genes.

(A) Exon-intron composition. (B) Expression patterns.

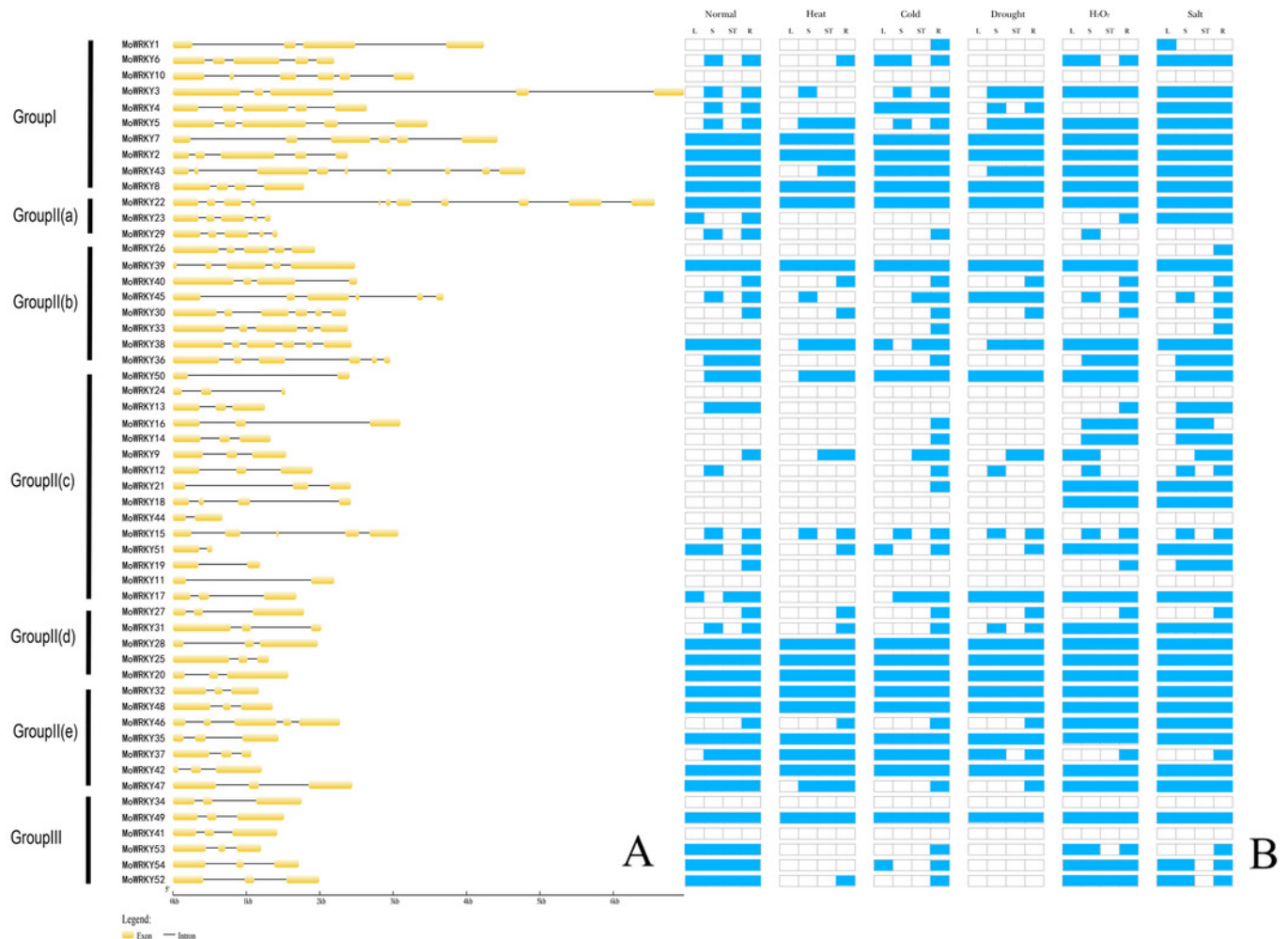



Figure 3

Distribution of conserved motifs in MoWRKYs.

Different colors represent different motifs.

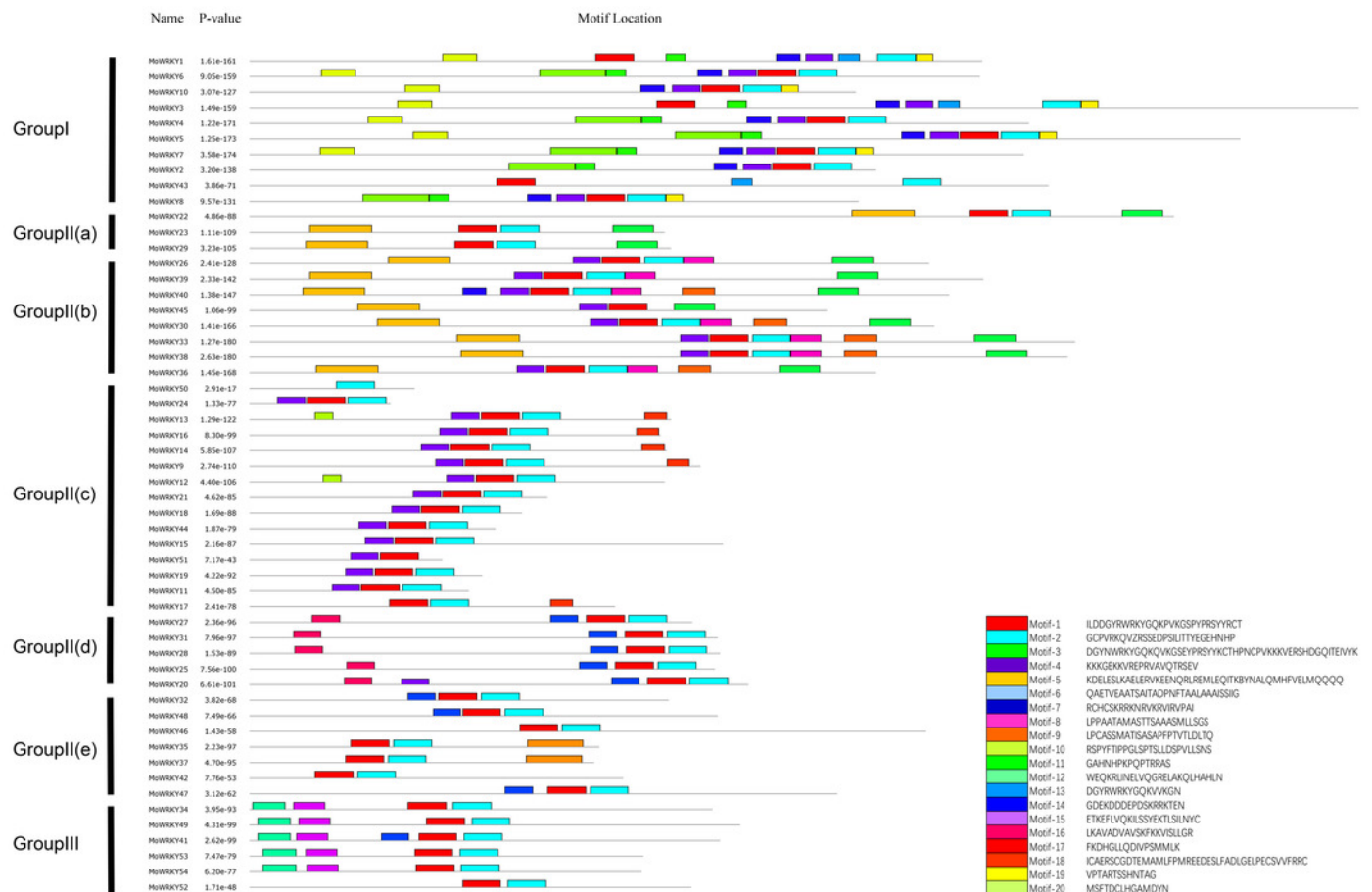



Figure 4

Phylogenetic tree of 81 group III WRKY proteins from drumstick and other seven species.

O. sativa (blue triangles and lines), *P. euphratica* (green lines), *V. vinifera* (orange lines), *P. patens* (purple triangle), *A. thaliana* (brown triangle and lines), *S. moellendorfii* (grass green lines) and *P. abies* (red lines).

MoWRKY52

MoWRKY34

MoWRKY41

MoWRKY49

MoWRKY53

MoWRKY54

0.07

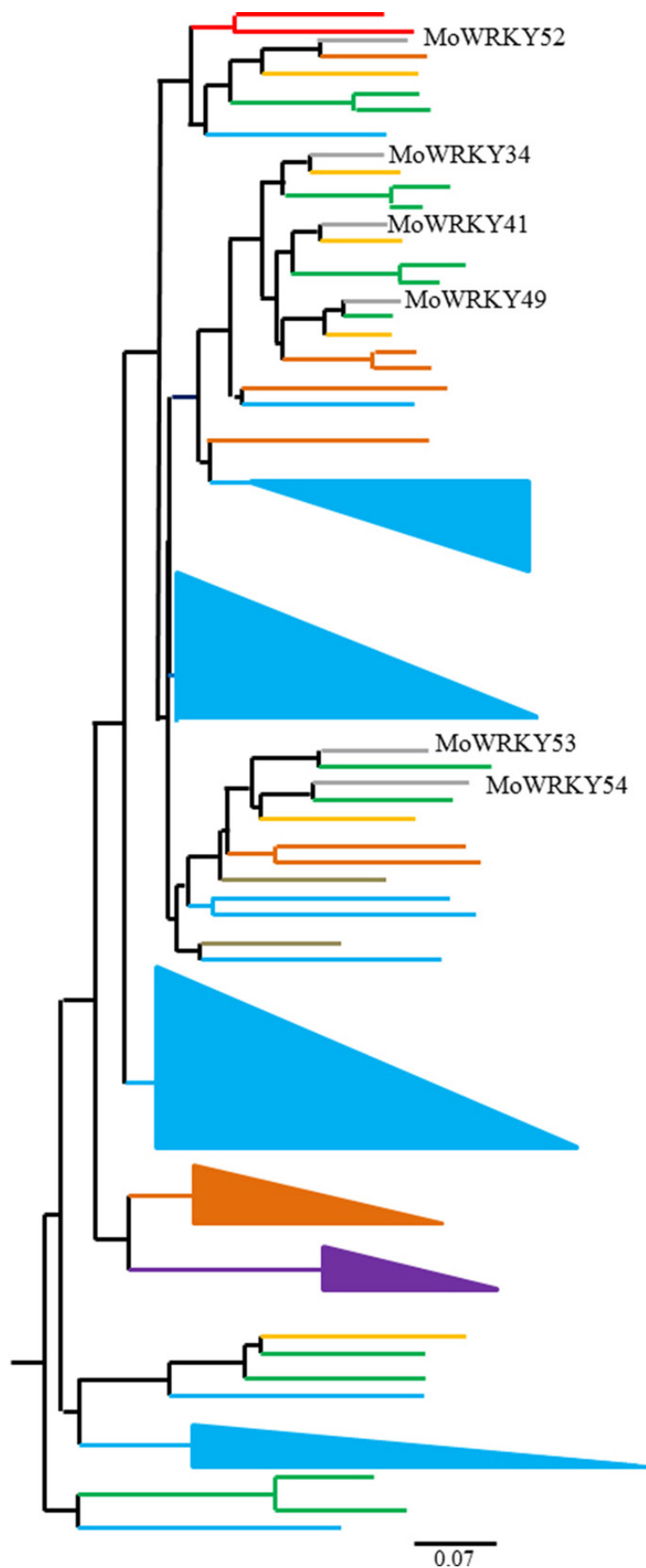



Figure 5

Frequency distribution and average values of Ka/Ks ratios.

(A) Frequency distribution between any two drumstick *WRKY* genes. (B) Average values of Ka/Ks across sub-groups of drumstick *WRKY*s.

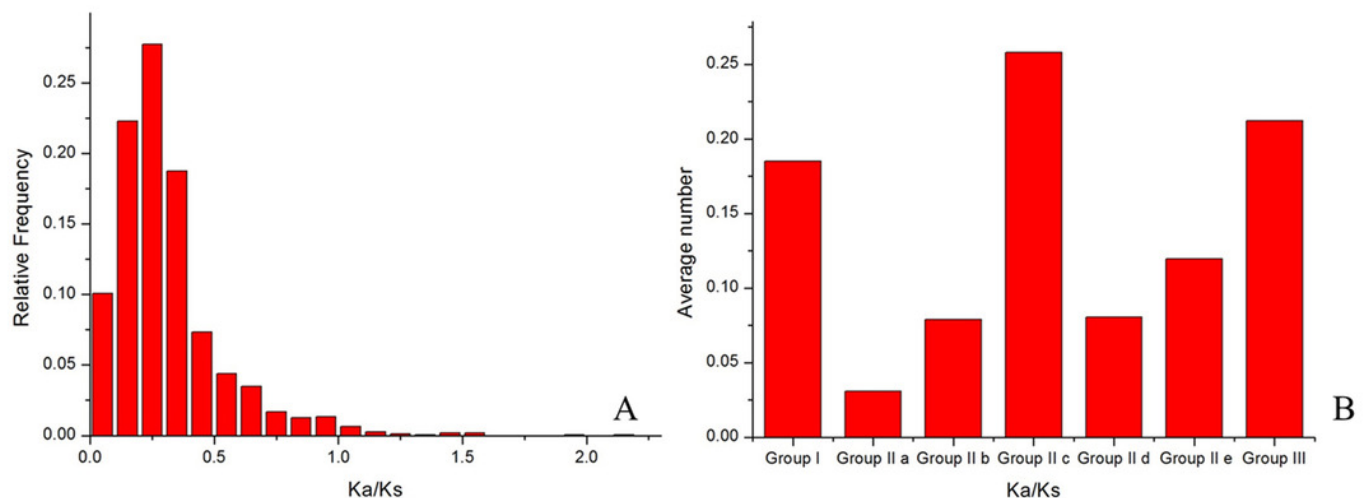
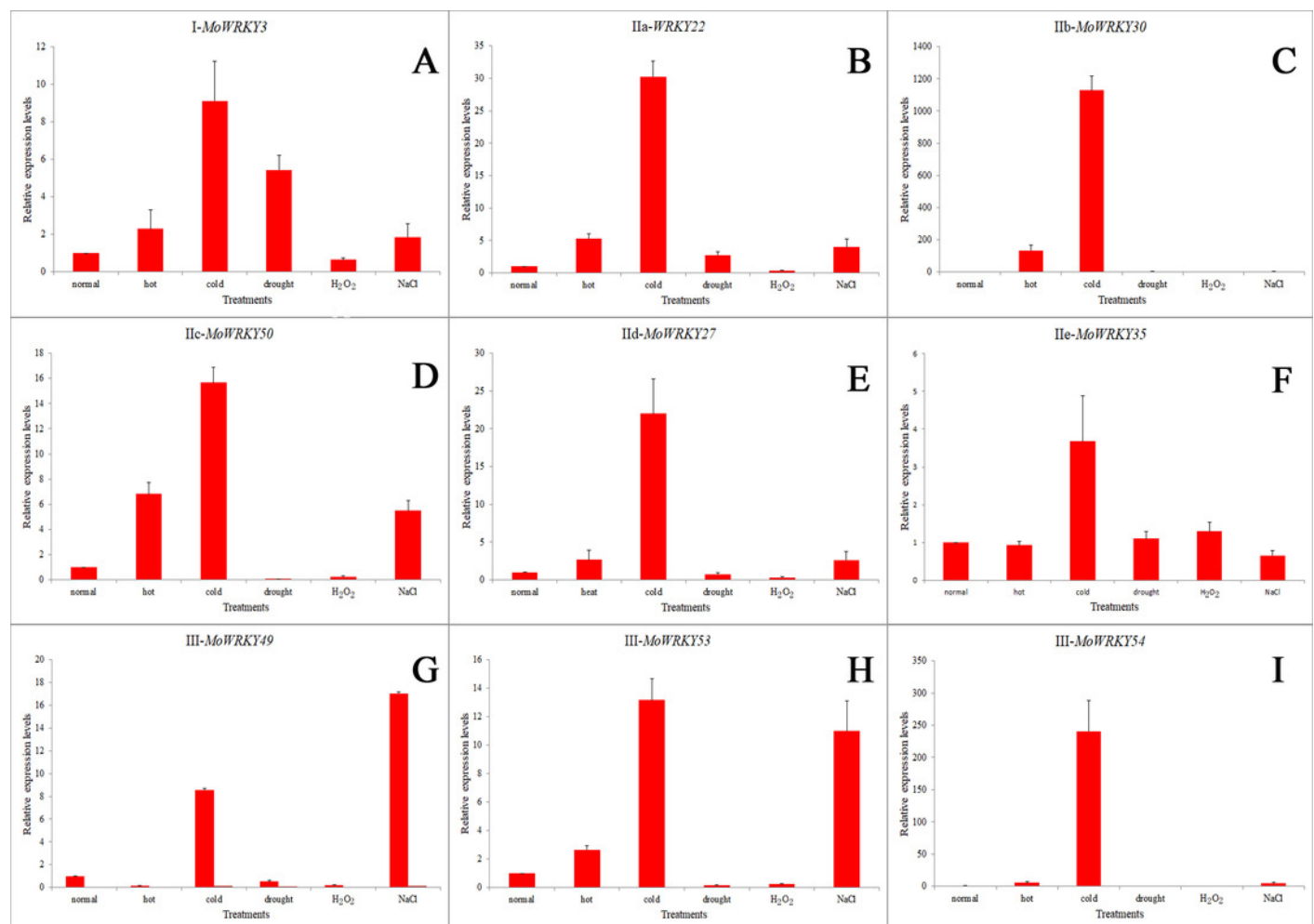



Figure 6

Expression profiles for 9 selected *MoWRKY* genes in root under different stresses.

(A) *I-MoWRKY3*; (B) *IIa-MoWRKY22*; (C) *IIb-MoWRKY30*; (D) *IIe-MoWRKY50*; (E) *IId-MoWRKY27*;

(F) *IIe-MoWRKY35*; (G) *III-MoWRKY49*; (H) *III-MoWKRY53*; (I) *III-MoWKRY54*.

Full-length WRKY genes identified from drumstick genome

1   Table 1 Full-length WRKY genes identified from drumstick genome

Class	Gene name	Annotation ID	Conserved motify	Zinc finger
I	*MoWRKY43*	lamu_GLEAN_10016673	WRKYGQK	$C\text{-}X_4\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY8*	lamu_GLEAN_10019070	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY2*	lamu_GLEAN_10014815	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY3*	lamu_GLEAN_10006432	WRKYGQK	$C\text{-}X_4\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY6*	lamu_GLEAN_10006277	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY4*	lamu_GLEAN_10010412	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY7*	lamu_GLEAN_10010176	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY5*	lamu_GLEAN_10005513	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY1*	lamu_GLEAN_10000767	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
I	*MoWRKY10*	lamu_GLEAN_10018171	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIa	*MoWRKY22*	lamu_GLEAN_10016899	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIa	*MoWRKY23*	lamu_GLEAN_10005532	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIa	*MoWRKY29*	lamu_GLEAN_10016902	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY26*	lamu_GLEAN_10015703	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY36*	lamu_GLEAN_10013925	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY30*	lamu_GLEAN_10010114	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY33*	lamu_GLEAN_10005737	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY38*	lamu_GLEAN_10016471	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY40*	lamu_GLEAN_10015347	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY39*	lamu_GLEAN_10018130	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIb	*MoWRKY45*	lamu_GLEAN_10004479	WRKYGQK	$C\text{-}X_5\text{-}C\text{-}X_{23}\text{-}HXH$
IIc	*MoWRKY17*	lamu_GLEAN_10015158	WRKYGQK	$C\text{-}X_4\text{-}C\text{-}X_{23}\text{-}HXH$
IIc	*MoWRKY21*	lamu_GLEAN_10005936	WRKYGQK	$C\text{-}X_4\text{-}C\text{-}X_{23}\text{-}HXH$
IIc	*MoWRKY18*	lamu_GLEAN_10014440	WRKYGQK	$C\text{-}X_4\text{-}C\text{-}X_{23}\text{-}HXH$
IIc	*MoWRKY16*	lamu_GLEAN_10002123	WRKYGQK	$C\text{-}X_4\text{-}C\text{-}X_{23}\text{-}HXH$

Class	Gene name	Annotation ID	Conserved motify	Zinc finger
IIc	*MoWRKY50*	lamu_GLEAN_10005926	-	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY9*	lamu_GLEAN_10018985	WRKYGQK	$C-X_4-C-X_{23}-HXH$
Class	Gene name	Annotation ID	Conserved motify	Zinc finger
IIc	*MoWRKY14*	lamu_GLEAN_10013856	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY24*	lamu_GLEAN_10017233	WRKYGKK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY13*	lamu_GLEAN_10016027	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY12*	lamu_GLEAN_10010840	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY44*	lamu_GLEAN_10009886	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY15*	lamu_GLEAN_10014128	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY51*	lamu_GLEAN_10003738	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY11*	lamu_GLEAN_10007141	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IIc	*MoWRKY19*	lamu_GLEAN_10017855	WRKYGQK	$C-X_4-C-X_{23}-HXH$
IId	*MoWRKY31*	lamu_GLEAN_10007564	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IId	*MoWRKY28*	lamu_GLEAN_10011212	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IId	*MoWRKY27*	lamu_GLEAN_10016840	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IId	*MoWRKY25*	lamu_GLEAN_10013546	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IId	*MoWRKY20*	lamu_GLEAN_10005795	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IIe	*MoWRKY47*	lamu_GLEAN_10007164	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IIe	*MoWRKY35*	lamu_GLEAN_10001324	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IIe	*MoWRKY37*	lamu_GLEAN_10016099	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IIe	*MoWRKY42*	lamu_GLEAN_10013842	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IIe	*MoWRKY46*	lamu_GLEAN_10012212	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IIe	*MoWRKY32*	lamu_GLEAN_10009888	WRKYGQK	$C-X_5-C-X_{23}-HXH$
IIe	*MoWRKY48*	lamu_GLEAN_10014133	WRKYGQK	$C-X_5-C-X_{23}-HXH$
III	*MoWRKY52*	lamu_GLEAN_10005191	WRKYGQK	$C-X_7-C-X_{23}-HXC$
III	*MoWRKY41*	lamu_GLEAN_10009829	WRKYGQK	$C-X_7-C-X_{23}-HXC$
III	*MoWRKY34*	lamu_GLEAN_10014082	WRKYGQK	$C-X_7-C-X_{23}-HXC$

III	*MoWRKY49*	lamu_GLEAN_10012174	WRKYGQK	$C-X_7-C-X_{23}-HXC$
III	*MoWRKY54*	lamu_GLEAN_10006335	WRKYGQK	$C-X_7-C-X_{23}-HXC$
III	*MoWRKY53*	lamu_GLEAN_10005192	WRKYGQK	$C-X_7-C-X_{23}-HXC$

The predicted stress-responsive *cis-acting* elements in the promoters of *MoWRKYs*

1   Table 2 The predicted stress-responsive *cis-acting* elements in the promoters of *MoWRKYs*

Cis-acting elements	Function	Genes
ABRE	Involved in ABA response	*MoWRKY1, 3, 4, 5, 6, 8, 9, 11, 13, 14, 15, 17, 19, 20, 21, 23, 24, 25, 26, 28, 29, 31, 33, 34, 37, 38, 40, 42, 43, 44, 46, 47, 48, 49, 51, 52, 53, 54*
ARE	Essential for the anaerobic induction	*MoWRKY 1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 15, 16, 17, 18, 20, 21, 23, 24, 26, 27, 28, 30, 31, 32, 33, 34, 37, 38, 41, 42, 43, 45, 46, 47, 49, 51, 52, 53, 54*
MBS	Involved in drought inducibility	*MoWRKY 2, 4, 6, 7, 8, 10, 11, 12, 13, 14, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 42, 43, 44, 45, 47, 48, 49, 50, 53, 54*
HSE	Involved in heat stress response	*MoWRKY 3, 4, 5, 9, 12, 14, 16, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 32, 33, 35, 36, 38, 39, 40, 41, 43, 45, 46, 47, 49, 50, 51, 53*
Sp1	Light responsive element	*MoWRKY 4, 5, 6, 9, 10, 11, 13, 14, 18, 19, 20, 21, 23, 24, 27, 28, 31, 35, 37, 38, 39, 41, 42, 44, 46, 47, 48, 50, 52, 53, 54*
G-box	ABA, light, UV and hurt responsive element	*MoWRKY 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 23, 24, 25, 26, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, 49, 51, 53, 54*
Box 4	Part of a conserved DNA module involved in light response	*MoWRKY 1, 2, 3, 4, 6, 8, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50, 51, 52, 53*
CGTCA motif	Involved in MeJA response	*MoWRKY 1, 3, 4, 6, 7, 8, 9, 11, 12, 15, 16, 17, 18, 19, 21, 22, 24, 25, 27, 28, 29, 30, 33, 34, 35, 37, 38, 39, 40, 41, 43, 44, 46, 48, 49, 50, 51, 52, 53, 54*
TGACG motif	Involved in MeJA response	*MoWRKY 1, 3, 4, 6, 7, 8, 11, 12, 15, 16, 17, 18, 19, 21, 22, 24, 25, 27, 28, 29, 30, 33, 34, 35, 37, 38, 39, 40, 41, 43, 44, 46, 48, 49, 50, 51, 52, 53, 54*
GT1 motif	Light responsive element	*MoWRKY 1, 6, 7, 9, 11, 12, 13, 15, 16, 18, 19, 20, 22, 23, 24, 26, 27, 29, 34, 35, 36, 37, 39, 40, 43, 44, 45, 47, 48, 49, 50, 54*