Genome-wide identification and characterization of WRKY gene family in *Salix suchowensis*

Changwei Bi [1] , **Yiqing Xu** [1] , **Qiaolin Ye** [1] , **Tongming Yin** [2] , **Ning Ye** Corresp. [1]

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing, Jiangsu, China

[2] College of Forest Resources and Environment, Nanjing Forestry University, Nanjing, Jiangsu, China

Corresponding Author: Ning Ye
Email address: yening@njfu.edu.cn

WRKY proteins are the zinc finger transcription factors that were first identified in plants. They can specifically interact with the W-box, which can be found in the promoter region of a large number of plant target genes, to regulate the expressions of downstream target genes. They also participate in diverse physiological and growing processes in plants. Prior to this study, a plenty of WRKY genes have been identified and characterized in herbaceous species, but there is no large-scale study of WRKY genes in willow . With the whole genome sequencing of Salix suchowensis, we have the opportunity to conduct the genome-wide research for willow WRKY gene family. In this study, we identified 85 WRKY genes in the willow genome and renamed them from SsWRKY1 to SsWRKY85 on the basis of their specific distributions on chromosomes. Due to their diverse structural features, the 85 willow WRKY genes could be further classified into three main groups (group I - III), with five subgroups (IIa - IIe) in group II. With the multiple sequence alignment and the manual search, we found three variations of the WRKYGQK heptapeptide: WRKYGRK, WKKYGQK and WRKYGKK, and four variations of the normal zinc finger motif, which might execute some new biological functions. In addition, the SsWRKY genes from the same subgroup share the similar exon–intron structures and conserved motif domains. Further studies of SsWRKY genes revealed that segmental duplication events played a more prominent role in the expansion of SsWRKY genes. Distinct expression profiles of SsWRKY genes with RNA sequencing data revealed that diverse expression patterns among five tissues, including tender roots, young leaves, vegetative buds, non-lignified stems and barks. With the analyses of WRKY gene family in willow, it is not only beneficial to complete the functional and annotation information of WRKY genes family in woody plants, but also provide important references to investigate the expansion and evolution of this gene family in flowering plants.

Genome-wide identification and characterization of WRKY

gene family in *Salix suchowensis*

Changwei Bi[1], Yiqing Xu[1], Yanshu Qu[1], Qiaolin Ye[1], Tongming Yin[2], Ning Ye[1*]

1. College of Information Science and Technology, Nanjing Forestry University, Nanjing,

Jiangsu, China

2. College of Forest Resources and Environment, Nanjing Forestry University, Nanjing, Jiangsu,

China

* Corresponding author:

Ning Ye[1]

Nanjing Forestry University, 159 Longpan Road, Nanjing, 210037, China

Email address: yening@njfu.edu.cn

12 **Abstract**

13 WRKY proteins are the zinc finger transcription factors that were first identified in plants. They

14 can specifically interact with the W-box, which can be found in the promoter region of a large

15 number of plant target genes, to regulate the expressions of downstream target genes. They also

16 participate in diverse physiological and growing processes in plants. Prior to this study, a plenty

17 of WRKY genes have been identified and characterized in herbaceous species, but there is no

18 large-scale study of WRKY genes in willow. With the whole genome sequencing of *Salix*

19 *suchowensis*, we have the opportunity to conduct the genome-wide research for willow WRKY

20 gene family. In this study, we identified 85 WRKY genes in the willow genome and renamed

21 them from SsWRKY1 to SsWRKY85 on the basis of their specific distributions on

22 chromosomes. Due to their diverse structural features, the 85 willow WRKY genes could be

23 further classified into three main groups (group I - III), with five subgroups (IIa - IIe) in group II.

24 With the multiple sequence alignment and the manual search, we found three variations of the

25 WRKYGQK heptapeptide: WRKYGRK, WKKYGQK and WRKYGKK, and four variations of

26 the normal zinc finger motif, which might execute some new biological functions. In addition, the

27 SsWRKY genes from the same subgroup share the similar exon–intron structures and conserved

28 motif domains. Further studies of SsWRKY genes revealed that segmental duplication events

29 played a more prominent role in the expansion of SsWRKY genes. Distinct expression profiles of

30 SsWRKY genes with RNA sequencing data revealed that diverse expression patterns among five

31 tissues, including tender roots, young leaves, vegetative buds, non-lignified stems and barks.

32 With the analyses of WRKY gene family in willow, it is not only beneficial to complete the

33 functional and annotation information of WRKY genes family in woody plants, but also provide

34 important references to investigate the expansion and evolution of this gene family in flowering

35 plants.

Introduction

37 Plants form a series of adjustment mechanisms to adapt diverse environment stress in their long

38 evolutionary processes. Among the numerous adjustment mechanisms, transcription factors play

39 important roles (Jang et al. 2010). In plants, WRKY proteins constitute a large family of

40 transcription factors, involving in various physiological and developmental processes (Eulgem

41 2000; Rushton et al. 2010). Since the first WRKY gene was cloned and characterized from sweet

42 potato (Ishiguro & Nakamura 1994), many corresponding studies have been conducted rapidly.

43 Such as *Arabidopsis thaliana*, desert legume (*Retama raetam*), cotton (*Gossypium arboreum*),

44 rice (*Oryza sativa*), *Pinus monticola*, barley (*Hordeum vulgare*), sunflower, cucumber (*Cucumis*

45 *sativus*), poplar (*Populus trichocarpa*), tomato (*Solanum lycopersicum*) and grapevine (*Vitis*

46 *vinifera*) (Ding et al. 2015; Eulgem 2000; Giacomelli et al. 2010; Guo et al. 2014; He et al. 2012;

47 Huang et al. 2012; Ling et al. 2011; Liu & Ekramoddoullah 2009; Mangelsen et al. 2008; Pnueli

48 et al. 2002; Wu 2005).

49 The existence of either one or two highly conserved WRKY domains is the most vital structural

50 characteristic of WRKY gene. WRKY gene consists of about 60 amino acid residues with a

51 conserved WRKYGQK heptapeptide at its N-termini, and a zinc finger motif (C-X_{4-5}-C-X_{22-23}-H-

52 X_1-H or C-X_7-C-X_{23}-H-X_1-C) at the C-terminal region. Previous functional studies indicated that

53 WRKY genes could specifically interact with the W-box ([C/T]TGAC[T/C]), the promoter

54 region of plant target genes, to adjust the expressions of downstream target genes (Ciolkowski et

55 al. 2008). Additionally, SURE (sugar responsive elements), another prominent *cis*-element that

56 can promote transcription processes, was also found to bind to the WRKY transcription factors

57 under a convincing research (Sun 2003). The proper DNA-binging ability of WRKY genes could

58 be influenced by the variation of the conserved WRKYGQK heptapeptide (Duan et al. 2007;

59 Maeo et al. 2001).

60   The WRKY proteins can be classified into three main groups (I, II and III) on the basis of the

61   number of their WRKY domains and the pattern of the zinc finger motif. Proteins from group I

62   contain two WRKY domains followed by a C_2H_2 zinc finger motif, while the other WRKY

63   proteins from group II and III only contain one WRKY domain followed by a C_2H_2 or C_2HC

64   correspondingly (Yamasaki et al. 2005). Group II can be further divided into five subgroups from

65   IIa to IIe based on additional amino acid motifs present outside the WRKY domain. Apart from

66   the conserved WRKY domains and the zinc finger motif, there are some WRKY proteins

67   appearing to have basic nuclear localization signal, LZs (leucine zipper) (Cormack et al. 2002),

68   serine-threonine-rich region, glutamine-rich region and proline-rich region (Ulker & Somssich

69   2004). Throughout the studies of WRKY gene family in many higher plants (Liu &

70   Ekramoddoullah 2009; Rushton et al. 2010; Wu 2005), WRKY genes have been identified to be

71   involved in various regulatory processes mediated by different biotic and abiotic stresses

72   (Ramamoorthy et al. 2008). In plant defense against various biotic stresses, such as bacterial,

73   fungal and viral pathogens, it has been well documented that the WRKY genes play vital roles

74   (Cheng et al. 2015; Dong et al. 2003; He et al. 2016; Jaffar et al. 2016; Jiang et al. 2016; Kim et

75   al. 2016; Li et al. 2006; Liu et al. 2016; Xu et al. 2006; Zhou et al. 2008). They are also involved

76   in abiotic stress-induced gene expression. In *Arabidopsis*, with the either heat or salt treatments,

77   the expressions of AtWRKY25 and AtWRKY33 are transformed apparently (Jiang & Deyholos

78   2009). Furthermore, the expression of TcWRKY53 that belonged to alpine penny grass (*Thlaspi*

79   *caerulescens*) is affected by salt, cold, and polyethylene glycol treatments (Wei et al. 2008). In

80   rice, a total of 54 OsWRKY genes showed noticeable differences in their transcript abundance

81   under the abiotic stress such as cold, drought, and salinity (Ramamoorthy et al. 2008). There is

82   also accumulating evidence that WRKY genes are involved in regulating developmental

83   processes, such as embryo morphogenesis (Lagace & Matton 2004), senescence (Robatzek &

84   Somssich 2002), trichome initiation (Johnson 2002), and some signal transduction processes

85   mediated by plant hormones including gibberellic acid (Zhang et al. 2004), abscisic acid (Zou et

86   al. 2004), or salicylic acid (Du & Chen 2008).

87   The number of WRKY genes in different species varies tremendously. For instance, there are 72

88   members in *Arabidopsis thaliana*, at least 45 in barley, 57 in cucumber, 58 in physic nut

89   (*Jatropha curcas*), 59 in grapevine, 104 in poplar, 105 in foxtail millet (*Setaria italica*), 112 in

90   *Gossypium raimondii* and more than 109 in rice (Ding et al. 2015; Eulgem 2000; Guo et al. 2014;

91   He et al. 2012; Ling et al. 2011; Mangelsen et al. 2008; Muthamilarasan et al. 2015; Wu 2005;

92   Xiong et al. 2013). Zhang et al. also identified the most basal WRKY genes in the lineage of non-

93   plant eukaryotes and green alga (Zhang & Wang 2005). Interestingly, the WRKY genes in

94   eukaryotic unicellular chlamydomonas, protoctist (*Giardia lambliad*), bryophyte (*Physcomitrella*

95   *patens*) and fern (*Ceratopteris richardii*) all belonged to group I (Di-Qiu et al. 2006; Ulker &

96   Somssich 2004; Zhang & Wang 2005). For example, the study in bryophyte (*Physcomitrella*

97   *patens*) found at least 12 WRKY genes, and all the genes belonged to group I (Ulker & Somssich

98   2004). Additionally, the study in gymnosperm (*Cycas revolute*) identified at least 21 WRKY

99   genes (Di-Qiu et al. 2006), and they were divided into two groups, 15 WRKY genes therein

100   belonged to group I and the other 6 WRKY genes belonged to group II. Further study suggested

101   that the core WRKY domains of group II and III were similar to the C-terminal domain of group

102   I, therefore, the group II WRKY genes might emerge from the breakage of the C-terminal domain

103   in group I and the group III probably evolve from group II (Ulker & Somssich 2004). Above of

104   all indicated that the group I WRKY genes might be the oldest type, which evolved from the

105   origin of eucaryon, and group II and III might generate after the origin of bryophyte (Xie et al.

106   2005; Zhang & Wang 2005). In the evolution of WRKY genes, gene duplication events played

107   prominent roles. As a matter of fact, gene duplication events can lead to the generation of new

108   genes. For example, there are approximately 80% of OsWRKY (rice) genes located in duplicated

109   regions (Wu 2005), as well as 83% of PtWRKY (poplar) genes (He et al. 2012). However, no

110   gene duplication events have occurred in cucumber (Ling et al. 2011).

111   In the last few years, the increasing consumption of fossil fuels induced in a substantial increase

112   of CO_2 concentration, which has adverse impacts on global climate changes(Pleguezuelo et al.

113   2014). Therefore, an ever-increasing demand for energy from renewable sources has provided a

114   new impetus to cultivate woody plants for bioenergy production. Due to its ease of propagation,

115   rapid growth and high yield on short rotation systems, some willow species have been used as

116   renewable resources since the 1970s. Additionally, with its essential physiological characteristics,

117   willow becomes prominent parts of the basket production, environmental restoration, analgesic

118   extraction, phytoremediation, both riparian and upland erosion control and biomass production

119   (Kuzovkina & Quigley 2005). WRKY proteins participate in diverse physiological and

120   developmental processes in plants. With these various important factors and the recent released

121   *Salix suchowensis* genome sequence, which covers about 96% of the expressed gene loci (Dai et

122   al. 2014), we have the opportunity to analyze the willow WRKY gene family. The

123   characterization of WRKY genes in willow can provide interesting gene pools to be investigated

124   for breeding and genetic engineering purposes in woody plants.

Materials and Methods

Datasets and sequence retrieval

127   The sequence of a shrub willow *Salix suchowensis* (*S. suchowensis*), which flowers within two

128   years, was conducted with a combined approach using Roche/454 and Illumina/HiSeq-2000

129   sequencing technologies (Dai et al. 2014). The latest v5.2 *S. suchowensis* genome annotation

130   information (version5_2.gff3) and protein sequences (Willow.gene.pep) were downloaded from

131   our laboratory website (http://bio.njfu.edu.cn/ss_wrky/). Sequences of 72 *Arabidopsis* WRKY

132   proteins were obtained from TAIR (release 10, http://www.arabidopsis.org/), and 104 poplar

133 WRKY proteins were obtained from the Supplementary material 3 of poplar (Eulgem 2000; He et

134 al. 2012).

Identification and distribution of WRKY genes in willow

136 The procedure performed to identify putative WRKY proteins in willow was similar to the

137 method described in other species (Guo et al. 2014; He et al. 2012; Wu 2005). The Hidden

138 Markov Model (HMM) profile for the WRKY transcription factor was downloaded from the

139 Pfam database (http://pfam.sanger.ac.uk/) with the keyword 'PF03106' (Punta et al. 2012). The

140 HMM profile was applied as a query to search against the all willow protein sequences

141 (Willow.gene.pep) using BLASTP program (E-value cutoff = $1e^{-3}$) (Camacho et al. 2009).

142 Another procedure was performed to validate the putative accuracy. An alignment of WRKY

143 seed sequences in Stockholm format from Pfam database was used by HMMER program

144 (hmmbuild) to build a HMM model, and then the model was used to search the willow protein

145 sequences by another HMMER program (hmmsearch) with default parameters (Eddy 1998).

146 Finally, we employed the SMART program (http://smart.embl-heidelberg.de/) to confirm the

147 candidates from the two procedures correlated with the WRKY structure features (Letunic et al.

148 2015).

149 Additionally, we calculated the length, MW (molecular weight), PI (isoelectric point) of these

150 putative WRKY proteins by ExPasy site (http://au.expasy.org/tools/pi_tool.html). Every WRKY

151 genes were mapped onto chromosomes (http://bio.njfu.edu.cn/ss_wrky/version5_2.fa) with an in-

152 house Perl script (http://bio.njfu.edu.cn/willow_chromosome/BuildGff3_Chr.pl), and then

153 renamed based on their orderly given chromosomal distribution. The distribution graph of every

154 WRKY gene was drawn by MapInspect software (http://mapinspect.software.informer.com/).

155 **Sequence alignments, phylogenetic analysis and classification of willow**

156 **WRKY genes**

157 Using the online tool SMART, we obtained the conserved WRKY core domains of predicted

158 SsWRKY genes, and then multiple sequence alignment based on these domains was performed

159 using ClustalX (version 2.1) (Larkin et al. 2007). After alignment, we used Boxshade

160 (http://www.ch.embnet.org/software/BOX_form.html) to color the alignment result online. To

161 gain better classification of these SsWRKY genes, a further multiple sequence alignment

162 including 103 SsWRKY domains and 82 WRKY domains from *Arabidopsis* (AtWRKY) was

163 performed using ClustalW (Larkin et al. 2007), and a phylogenetic tree based on this alignment

164 was built by MEGA 6.0 with the Neighbor-joining (NJ) method (Tamura et al. 2013). Bootstrap

165 values have been calculated from 1000 iterations in the pairwise gap deletion mode, which is

166 conducive to the topology of the NJ tree by divergent sequences. Based on the phylogenetic tree

167 constructed by SsWRKY and AtWRKY domains, these SsWRKY genes were classified into

168 different groups and subgroups. In order to get a better comparison of WRKY family in

169 Salicaceae, a phylogenetic tree including all SsWRKY domains and 126 WRKY domains from

170 poplar (PtWRKY) was constructed with the similar method to *Arabidopsis*. Additionally, a

171 phylogenetic tree based on full-length SsWRKY genes was also constructed to get a better

172 classification. The ortholog of each SsWRKY gene in *Arabidopsis* and poplar was based on the

173 phylogenetic trees of their respective WRKY domains, and the members of group I WRKY genes

174 were considered as orthologs unless the same phylogenetic relationship can be detected between

175 N-termini and C-termini in the tree. Another method described by Zou et al., BLAST-based

176 method (Bi-direction best hit), was used to verify the putative orthologous genes (E-value cutoff

177 = 1e-20) (Chen et al. 2007; Zou et al. 2016).

Evolutionary analysis of WRKY III genes in willow

The group of WRKY III genes, only found in flowering plants, is considered as the evolutionary youngest groups, and plays crucial roles in the process of plant growth (He et al. 2012; Wu 2005). As described by Wang et al., the WRKY III genes also have a prominent impact on disease and drought resistance (Wang et al. 2015). Previous study of Zhang et al. held the opinion that duplications and diversifications were plentiful in WRKY III genes, and they appeared to have confronted different selection challenges (Zhang & Wang 2005). Phylogenetic analysis of WRKY III genes was performed using MEGA6.0 with 65 WRKY III genes from *Arabidopsis* (AtWRKY), *Populus* (PtWRKY), grape (VvWRKY), willow (SsWRKY) and rice (OsWRKY). A NJ tree was constructed with the same method described before. Additionally, we estimated the non-synonymous (Ka) and synonymous (Ks) substitution ratio of SsWRKY III genes to verify whether selection pressure participated in the expansion of SsWRKY III genes. Each pair of these WRKY III protein sequences was first aligned using ClustalW. The alignments generated by ClustalW and the corresponding cDNA sequences were submitted to the online program PAL2NAL (http://www.bork.embl.de/pal2nal/) (Suyama et al. 2006), which automatically calculates Ks and Ka by the codeml program in PAML (Yang 2007).

Analysis of exon-intron structure, gene clusters, gene duplication events and

conserved motif distribution of willow WRKY genes

The exon-intron structures of the willow WRKY genes were obtained based on the protein annotation files assembled ourselves (http://bio.njfu.edu.cn/ss_wrky/version5_2.gff3), and the diagrams were obtained from the online website Gene Structure Display Server (GSDS: http://gsds.cbi.pku.edu.cn/) (Hu et al. 2015).

200   Gene clusters, defined as a single chromosome containing two or more genes within 200 kb (He

201   et al. 2012; Holub 2001), are very important for predicting co-expression genes or potential

202   function of clustered genes in angiosperms (Overbeek et al. 1999).

203   Gene duplication events were always considered as the vital sources of biological evolution.

204   Two or more adjacent homologous genes located on a single chromosome were considered as

205   tandem duplication events (TDs), while homologous gene pairs between different chromosomes

206   were defined as segmental duplication events (SDs) (Liu & Ekramoddoullah 2009). BLASTP (E-

207   value cutoff = 1e-20) was performed to identify the gene duplication events in SsWRKY genes

208   with the following definition (Gu et al. 2002; He et al. 2012): (1) the coverage of the aligned

209   sequence ≥80% of the longer gene; and (2) the similarity of the aligned regions ≥70%. In this

210   study, we set the cutoff of the similarity of the aligned regions as 65%, because the similarity of

211   the unaligned regions may reduce the value in different species.

212   To better exhibit the structural features of SsWRKY proteins, the online tool MEME (Multiple

213   Expectation Maximization for Motif Elicitation) was used to identify the conserved motifs in the

214   encoded SsWRKY proteins (Bailey et al. 2006). The optimized parameters were employed as the

215   following: any number of repetitions, maximum number of motifs = 20, and the optimum width

216   of each motif was constrained to between 6 to 50 residues. The online program 2ZIP

217   (http://2zip.molgen.mpg.de/) was used to verify the existence of the conserved Leu zipper motif

218   (Bornberg-Bauer 1998), whereas some other important conserved motifs, HARF, LXXLL (X,

219   any amino acid) and LXLXLX, were identified manually.

220   **Expression analyses of willow WRKY genes**

221   The sequenced *S. suchowensis* RNA-HiSeq reads from five tissues including tender roots, young

222   leaves, vegetative buds, non-lignified stems and barks generated in our previous study were

223   separately mapped back onto the SsWRKY gene sequences using BWA (mismatch ≤ 2bp, other

224   parameters as default) (Li & Durbin 2009), and the number of mapped reads for each WRKY

225   gene was counted. Normalization of the mapped reads was done using RPKM (reads per kilo

226   base per million reads) method (Wagner et al. 2012). The heat map for tissue-specific expression

227   profiling was generated based on the $\log_2$RPKM values for each gene in all the tissue samples

228   using R package (Gentleman et al. 2004).

229   **Results**

230   **Identification and characterization of 85 WRKY genes in willow (*Salix***

231   *suchowensis*)**

232   In this study, we obtained 92 putative WRKY genes by using HMMER to search the Hidden

233   Markov Model profile of WRKY DNA-binding domain against willow protein sequences, and

234   validated the accuracy of the consequence by BLASTP. After submitting the 92 putative WRKY

235   genes to the online program SMART, seven genes without a complete WRKY domain were

236   removed, while the other 85 WRKY genes were selected as possible members of the WRKY

237   superfamily.

238   WRKY genes contain one or two WRKY domains, comprising a conserved WRKYGQK

239   heptapeptide at the N-termini and a novel zinc finger motif (C-X_{4-7}-C-X_{22-23}H-X-H/C) at the C-

240   termini (Eulgem 2000). The variations of WRKY core domain or zinc finger motif may lead to

241   the binding specificities of WRKY genes, but this remains to be largely demonstrated (Brand et

242   al. 2013; Rinerson et al. 2015; Yamasaki et al. 2005). In order to identify the variations in WRKY

243   core domains, a multiple sequence alignment of 85 SsWRKY core domains was conducted, and

244   the result was shown in Fig. 1. Among the selected 85 WRKY genes, 81 (95.3%) were identified

245   to have highly conserved sequence WRKYGQK, whereas the other four WRKY genes

246   (SsWRKY14, SsWRKY23, SsWRKY38 and SsWRKY78) had a single mismatched amino acid

247   in their core WRKY domains (Fig. 1). In SsWRKY14 and SsWRKY38, the WRKY domain has

248   the sequence WRKYGKK, while SsWRKY23 contains a WKKYGQK sequence, and

249   SsWRKY78 contains WRKYGRK sequence. Eulgem et al. previously described that the zinc

250   finger motif ($C-X_{4-5}-X_{22-23}-H-X_1-H$ or $C-X_7-C-X_{23}-H-X_1-C$) is another vital feature of the WRKY

251   family (Eulgem 2000). As illustrated in Fig. 1, four WRKY domains (SsWRKY76C,

252   SsWRKY64, SsWRKY12 and SsWRKY28) do not contain any distinct zinc finger motif, but

253   they were still reserved in the succeeding analyses, as performed in barley and poplar (He et al.

254   2012; Mangelsen et al. 2008). Additionally, some zinc-finger-like motifs, including $C-X_4-C-X_{21}-$

255   $H-X_1-H$ in SsWRKY23 and $C-X_5-C-X_{19}-H-X_1-H$ in SsWRKY73 and SsWRKY17, were

256   identified in willow WRKY genes. Both the two zinc-finger-like motifs were also found in poplar

257   (PtWRKY39, 57, 42 and 53).

258   Detailed characteristics of SsWRKY genes are listed in Table 1, including the WRKY gene

259   specific group numbers, chromosomal distribution, *Arabidopsis* and poplar orthologs. The

260   molecular weight (MW), isoelectric point (PI) and the length of each WRKY protein sequence

261   are also shown in Table 1. According to the particularization (Table 1), the average length of

262   these protein sequences is 407 residues, and the lengths ranged from 109 residues (SsWRKY23)

263   to 1,593 residues (SsWRKY78). Additionally, the isoelectric point (PI) ranged from 5.03

264   (SsWRKY38, SsWRKY60) to 10.27 (SsWRKY28), and the molecular weight (MW) ranged from

265   12.9 (SsWRKY23) to 179.0 kDa (SsWRKY78).

266   **Locations and gene clusters of willow WRKY genes**

267   84 of the 85 putative SsWRKY genes could be mapped onto 19 willow chromosomes and then

268   renamed from SsWRKY1 to SsWRKY84 based on their specific distributions on chromosomes.

269   Only one SsWRKY gene (willow_GLEAN_10002834), renamed as SsWRKY85, could not be

270   conclusively mapped onto any chromosome. As shown in Fig. 2, Chromosome (Chr) 2 possessed

271   the largest number of SsWRKY genes (11 genes), followed by Chr14 (10 genes). Eight SsWRKY

272   genes were found on Chr6, six on Chr1 and Chr16, and five on Chr5. Additionally, four

273   chromosomes (Chr4, Chr11, Chr17, Chr18) had four SsWRKY genes, as well as three SsWRKY

274   genes were found on Chr8, Chr13 and Chr19. Chr10 and Chr15 had two SsWRKY genes, and

275   only one SsWRKY gene was identified on Chr7, Chr9 and Chr12. The distribution of each

276   SsWRKY genes was extremely irregular, indicating the reduction of the tandem duplication

277   events in willow WRKY genes.

278   As described by Holub et al. (He et al. 2012; Holub 2001), a single chromosome region

279   containing two or more genes within 200 kb was defined as gene clusters, which are very

280   important for predicting co-expression genes or potential function of clustered genes in

281   angiosperms (Overbeek et al. 1999). According to this description, a total of 23 SsWRKY genes

282   were clustered into 11 clusters in willow (Fig. 2). The chromosomal distribution of gene cluster

283   was irregular, and only seven chromosomes were identified to have gene clusters. Three clusters,

284   including seven SsWRKY genes, were found on Chr2, and two clusters were found on both Chr6

285   and Chr14. Only one cluster was distributed on each of Chr3, Chr8, Chr10 and Chr18, whereas

286   none was identified on other eleven chromosomes. Further analysis of SsWRKY chromosomal

287   distribution showed that a high WRKY gene density region in only 2.23 Mb regions on Chr2,

288   which had also been observed in rice and poplar (He et al. 2012; Wu 2005).

Phylogenetic analysis and classification of WRKY genes in willow

290   In order to get a better separation of different groups and subgroups in SsWRKY genes, a total of

291   185 WRKY domains, including 82 AtWRKY domains and 103 SsWRKY domains, were used to

292   construct the NJ phylogenetic tree. On the basis of the phylogenetic tree and structural features of

293   WRKY domains, all 85 SsWRKY genes were clustered into three main groups (Fig. 3). Nineteen

294   members containing two WRKY domains and C_2H_2-type zinc finger motifs were categorized into

295   group I, except SsWRKY78, which contains only one WRKY domain and two zinc finger motifs.

296   Domain acquisition and loss events appear to have shaped the WRKY family (Ross et al. 2007;

297   Rossberg et al. 2001). Thus, SsWRKY78 may have evolved from a two-domain WRKY gene but

298   lost one WRKY domain during evolution. Additionally, as shown in Fig. 3, SsWRKY78 shows

299   high similarities to SsWRKY40N, implying a common origin of their domains. The similar

300   phenomenon was also found in PtWRKY90 of poplar (He et al. 2012).

301   The largest number of SsWRKY genes, comprising a single WRKY domain and C_2H_2 zinc

302   finger motif, were categorized into group II. SsWRKY genes of group II could be further divided

303   into five subgroups: IIa, IIb, IIc, IId and IIe. As shown in Fig. 3, subgroup IIa (4 members) and

304   IIb (8 members) were clustered into one clade, as well as subgroup IId (13 members) and IIe (11

305   members). Strikingly, SsWRKY genes in subgroup IIc (21 members) and group IC are classified

306   into one clade, suggesting that group II genes are not monophyletic and the group IIc WRKY

307   genes may evolve from the group I genes by the loss of the WRKY domain in N-terminal. As

308   shown in Fig. 3 and Fig. S1, SsWRKY23, SsWRKY34 and their orthologous genes

309   (AtWRKY49, PtWRKY39, PtWRKY57, PtWRKY34 and PtWRKY32) seem to form a new

310   subgroup closer to the group III. However, SsWRKY23 and SsWRKY34 exhibit the zinc finger

311   motif $C-X_4-C-X_{21}-H-X-H$ and $C-X_4-C-X_{23}-H-X-H$ as observed in the subgroup IIc and group IC.

312   Thereby, they were classified into subgroup IIc in this study.

313   Different from the C_2H_2 zinc finger pattern in group I and II, group III WRKY genes (7

314   members), broadly considered as playing vital roles in plant evolution process and adaptability,

315   contained one WRKY domain and a $C-X_7-C-X_{23}-H-X-C$ zinc finger motif. However, in rice and

316   barley, a new $CX_7CX_nHX_1C$ ($n \geq 24$) zinc finger motif was identified in group III (Mangelsen et

317   al. 2008; Wu 2005), which was never found in poplar, grape, *Arabidopsis* and willow, suggesting

318   that this feature perhaps only belong to monocotyledonous species.

319   In order to obtain a better study in woody plant species, a phylogenetic tree based on the WRKY

320   domains between willow and poplar was constructed (Fig. S1). The tree showed that most of the

321   WRKY domains from willow and poplar were clustered into sister pairs, suggesting that gene

322   duplication events played prominent roles in the evolution and expansion of WRKY gene family.

323   Furthermore, a total of twenty SsWRKY domains show extremely the same domains (similarity:

324   100%) to poplar, i.e., SsWRKY39 and PtWRKY9, SsWRKY39 and PtWRKY9, SsWRKY39 and

325   PtWRKY9, SsWRKY39 and PtWRKY9, and so on. Further functional analyses of these genes in

326   willow or poplar will provide a useful reference for another one.

327   **The ortholog of SsWRKY genes in *Arabidopsis* and poplar**

328   The clustering of orthologous genes emphasizes the conservation and divergence of gene

329   families, and they may contain the same functions (Ling et al. 2011). In this study, a phylogeny-

330   based method was used to identify the putative orthologous SsWRKY genes in *Arabidopsis* and

331   poplar (Fig. 3 and Fig. S1), and BLAST-based method (Bi-direction best hit) was used to confirm

332   the true orthologs. The WRKY genes of group I contained two WRKY domains, and both of

333   them were used to construct the phylogenetic trees. To avoid the mistakes of orthologous genes

334   in group I, the members of group I WRKY genes were considered as orthologous genes unless

335   the same phylogenetic relationship can be detected between N-termini and C-termini in the

336   phylogenetic tree. For example, SsWRKY37 and AtWRKY44 were considered as an orthologous

337   gene pair because they clustered into a clade of their N-termini and C-termini (Fig. 3), while

338   SsWRKY80 and PtWRKY30 were excluded from orthologous gene pairs due to their different

339   clusters of N-termini and C-termini (Fig. S1). Totally, 75 orthologous gene pairs were found

340   between willow and *Arabidopsis*, less than 82 orthologous gene pairs between willow and poplar

341   (Table 1), which was congruent with the evolutionary relationship among the three plant species.

Evolutionary analysis of WRKY III genes in willow

342

343 The WRKY III genes were considered as the evolutionary youngest groups, and played crucial

344 roles in the process of plant growth and resistance. In order to further probe the duplication and

345 diversification of WRKY III genes after the divergence of the monocots and dicots, a

346 phylogenetic tree was constructed using 65 WRKY III genes from *Arabidopsis* (13), rice (29),

347 poplar (10), willow (7) and grape (6). As shown in Fig. S2, willow SsWRKY III genes were

348 closer to the eurosids I group (poplar and grape) than eurosids II group (*Arabidopsis*) and

349 monocots (rice). Meanwhile, most *Arabidopsis* and rice WRKY III genes formed the relatively

350 independent clades, suggesting that two gene duplication events, including tandem and segmental

351 duplication, perhaps were the main factors in the expansion of WRKY III genes in *Arabidopsis*

352 and rice. The results also indicated that WRKY III genes might arise after the divergence of the

353 Arabidopsis (eurosids I) and eurosids II (poplar, willow and grape). The study by Ling et al. in

354 cucumber (Ling et al. 2011) showed the similar results and hence proved the validity.

355 Additionally, we found that seven rice WRKY III genes (OsWRKY55, 84, 18, 52, 46, 114 and

356 97) contained the variant domain WRKYGEK, which was not found in other four dicots

357 (*Arabidopsis*, poplar, grape and willow), implying that this may be a feature of WRKY III genes

358 in monocots and these OsWRKY genes may respond to different environmental signals.

359 According to the comparison of the number of WRKY III genes in the five observed plants, the

360 number is smaller in eurosids I (poplar, grape and willow) than *Arabidopsis* (eurosids II) and rice

361 (monocots), which may be caused by different patterns of duplication events. Genes generated by

362 duplication events are not stable, and can be retained or lost due to different selection pressure

363 and evolution (Zhang 2003). In order to determine which selection pressure played prominent

364 roles in the expansion of willow WRKY III genes, we estimated the Ka/Ks ratios for all pairs (21

365 pairs) of willow WRKY III genes. As shown in Table S1, all the Ka/Ks ratios were less than 0.5,

366   suggesting willow WRKY III genes had mainly been subjected to strong purifying selection and

367   they were slowing evolving at the protein level.

368   **Exon–intron structures of SsWRKY genes**

369   The exon-intron structures of multiple gene families play crucial roles during plant evolution. As

370   shown in Fig. 4, the SsWRKY gene phylogenetic tree and the corresponding exon-intron

371   structures are shown in A and B, respectively. Exon-intron structures of each group were shown

372   in Fig. 4B, a large number of WRKY genes had two to five introns (94%, 80 of 85), including 8

373   WRKY genes contained one intron; 39 contained two introns; 13 contained three introns; 15

374   contained four introns and 5 contained five introns. The number of exons in remaining WRKY

375   genes was quite different: SsWRKY49, SsWRKY76 and SsWRKY78 had six, eleven and ten

376   introns, respectively; SsWRKY17 had the largest number of introns (seventeen introns), while no

377   intron was found in SsWRKY12. The intron acquisition or loss occurred during the evolution of

378   WRKY gene family, while WRKY genes in the same group shared the similar number of introns

379   (Guo et al. 2014). In our study, most of WRKY genes in group I had three to six introns, expect

380   SsWRKY76 and SsWRKY78, which might acquire some introns during evolution. The number

381   of introns of WRKY genes in group II was extremely different, ranging from one to five introns,

382   except SsWRKY17 with 17 introns and SsWRKY12 with zero intron might obtain or loss some

383   introns during evolution. Strikingly, WRKY genes in group III had the most stable number of

384   introns with all of seven WRKY III genes had two introns, suggesting that WRKY III genes may

385   be the most stable genes in the environmental stress. The stable number of introns in SsWRKY

386   III genes was consistent with the results of Ka/Ks analysis, which reflected that purifying

387   selection pressure played vital roles in willow WRKY III genes.

388   A great deal of studies in WRKY genes proved that nearly all of the WRKY genes contained an

389   intron in their WRKY core domains (Eulgem 2000; Guo et al. 2014; He et al. 2012; Huang et al.

390   2012; Ling et al. 2011; Zou et al. 2004). According to the further analysis of SsWRKY genes,

391   two major types of splicing introns, R-type and V-type, introns were observed in numerous

392   SsWRKY domains. The R-type intron was spliced exactly at the R residue, about five amino

393   acids before the first Cys residue in the C_2H_2 zinc finger motif. The V-type intron was localized

394   before the V residue, six amino acids after the second Cys residue in the C_2H_2 zinc finger motif.

395   As shown in Fig. 4B, the R-type introns could be observed in more groups, including group IC,

396   subgroup IIc, IId, IIe and group III, while V-type introns were only observed in subgroup IIa and

397   IIb. However, there was no intron found in group IN. The similar results were also observed in

398   *Arabidopsis*, poplar and rice, suggesting that the special distribution of introns in WRKY

399   domains was a feature of WRKY family (Eulgem 2000; He et al. 2012; Wu 2005).

Identification of gene duplication events and conserved motifs in willow

401   Gene duplication events were always considered as the vital sources of biological evolution

402   (Chothia et al. 2003; Ohno et al. 2009). TDs were defined as two or more adjacent homologous

403   genes located on a single chromosome, while homologous gene pairs between different

404   chromosomes were defined as SDs (Liu & Ekramoddoullah 2009). In our study, a total of 33

405   homologous gene pairs, including 66 SsWRKY genes, were identified to participate in gene

406   duplication events (Table S3). The composition of gene duplication events in each group in

407   ascending order was group I: 73.7% (14 of 19), group II: 78% (46 of 59) and group III: 85.7% (6

408   of 7). Among the 33 homologous gene pairs, none of them appeared to have undergone TDs, on

409   the contrary, all of the 66 genes (77.6% of all SsWRKY genes) participated in SDs, implying that

410   segmental duplication events played major roles in the expansion of willow WRKY genes.

411   WRKY genes shared more functional and homologies in their conserved WRKY core domains

412   (about 60 residues), while the rest sequences of WRKY genes shared a little (Eulgem 2000). In

413   order to get a more comprehensive understanding of the structural feature in WRKY domains, the

414   conserved motifs of SsWRKY genes were predicted using the online program MEME (Fig. S3

415   and Table S2). Among the 20 putative motifs, motifs 1, 2, 3 and 5, broadly distributed across

416   SsWRKY genes, were characterized as the WRKY conserved domains. The motif 6 was

417   characterized as nuclear localization signals (NLS), which mainly distributed in subgroup II d

418   and IIe and group III. Some other motifs with poorly defined recently were also predicted by

419   MEME: the motif 4 was only found in group IC and subgroup IIc; motifs 7 and 9 were limited to

420   subgroup IIa and IIb; the motif 8 was found in group I and a few genes of subgroup IIc; motifs

421   10, 13, 15 and 17 were unique in subgroup IId; the motif 12 was only observed in subgroup IIb;

422   the motif 16 was mainly found in group II; the motif 18 was found in subgroup IIc; motifs 19 and

423   20 were only observed in subgroup I. The distinct conserved motifs of different groups could be

424   an important foundation for future structural and functional study in WRKY gene family.

425   Some other important motifs, including Leu zipper motif, HARF, LXXLL and LXLXLX, could

426   be also identified in WRKY genes. Using the online program 2ZIP, the conserved Leu zipper

427   motif, described as a common hypothetical structure to DNA binding proteins (McInerney et al.

428   1998), was identified in only two SsWRKY genes (SsWRKY61 and SsWRKY39). With manual

429   inspection, the conserved HARF (RTGHARFRR[A/G]P) motifs, whose putative functions were

430   not distinguished clearly, were only observed in seven WRKY genes of subgroup IId, including

431   SsWRKY82, 33, 45, 81, 9, 30 and 56. In the meantime, the conserved LXXLL and LXLXLX (L:

432   Leucine; X: any amino acid) motifs, which respectively defined as the co-activator and active

433   repressor motifs, were also found in SsWRKY genes. A total of seven SsWRKY genes

434   (SsWRKY19, 45, 72, 61, 76, 30 and 59) contained the helical motif LXXLL, whereas eight genes

435   (SsWRKY66, 26, 35, 81, 83, 75, 73 and 3) shared the LXLXLX motif. The plenty of conserved

436   motifs in WRKY genes with different lengths and variant functions, suggesting that the WRKY

437   genes might play more vital roles in gene regulatory network.

438 **Distinct expression profiles of SsWRKY genes in various tissues**

439 In order to gain more information about the roles of WRKY genes in willow, RNA-seq data from

440 the sequenced genotype were used to quantify the expression level of WRKY genes in five

441 tissues of *Salix suchowensis*. As illustrated in Fig. 5, the expression of all 85 SsWRKY genes

442 were detected in at least one of the five examined tissues, such as 84 genes in roots, 80 in stems,

443 84 in barks, all in buds and 73 in leaves. Meanwhile, the cluster analysis of the expression pattern

444 in five tissues showed that SsWRKY genes shared more similarities between stem and leaf, as

445 well as bark and bud, and root was more similar to the clade formed by bark and bud. The results

446 detected here were consistent with their biological characteristics. SsWRKY38, not detected in

447 roots and leaves, was also lowly expressed in other tissues. Similarly, SsWRKY74, not detected

448 in stems, barks and leaves, was only expressed in roots and buds with extremely low levels.

449 Among the five genes not expressed in stems, SsWRKY66, 74 and 79 were also not detected in

450 leaves. The largest number of expressed or unexpressed SsWRKY genes (12 genes) was found in

451 buds or leaves, respectively, suggesting that WRKY genes might play more roles in buds than

452 leaves.

453 According to the expression annotation of 85 SsWRKY genes by RPKM method in Fig. 5 and

454 Table S4, the total transcript abundance of SsWRKY genes in tender root (RPKM = 1181.21),

455 bark (RPKM = 1363.01) and vegetative bud (RPKM = 928.58) was relatively larger than that in

456 other two tissues, including non-lignified stem (RPKM = 537.88) and young leaf (RPKM =

457 349.84). As shown in Table S4, SsWRKY81 (RPKM = 97.75), the most expressed SsWRKY

458 genes in roots, was also expressed in other four tissues, though the expression levels were

459 relatively low; SsWRKY56 (RPKM = 32.54), the most expressed SsWRKY genes in stem, was

460 also highly expressed in other examined tissues. Similarly, SsWRKY67, the most expressed

461 SsWRKY genes in barks (RPKM = 188.16), was also detected in vegetative buds (RPKM

462 =82.07) and young leaves (RPKM = 26.11) with high expression levels. Similarly, SsWRKY6

463   (RPKM = 26.31), the most expressed genes in leaves, was also highly detected in other tissues. A

464   few genes, i.e., SsWRKY52, SsWRKY2 and SsWRKY35, were expressed highly in barks, but

465   lowly in other four tissues. The results mentioned above may be an important foundation for the

466   specific expression analysis of each WRKY gene in willow.

467   Discussion

468   The WRKY transcription factor gene family can specifically interact with the W-box to regulate

469   the expressions of downstream target genes. They also play prominent roles in diverse

470   physiological and growing processes, especially in various abiotic and biotic stress responses in

471   plants. Previous studies about the features and functions of WRKY family have been conducted

472   in many model plants, including *Arabidopsis* for annual herbaceous dicots (Eulgem 2000), grape

473   for perennial dicots (Guo et al. 2014), poplar for woody plants and rice for monocots (He et al.

474   2012; Wu 2005), but there is no large-scale study of WRKY genes in willow. Here, the

475   comprehensive analysis of WRKY family in willow (*Salix suchowensis*) would facilitate a better

476   understanding of WRKY gene superfamily and provide interesting gene pools to be investigated

477   for breeding and genetic engineering purposes in woody plants.

478   As described in many previous studies, the presence of highly conserved WRKY domains in

479   WRKY proteins is the most prominent characteristic of the WRKY gene family (Ding et al.

480   2015; Eulgem 2000; He et al. 2012; Huang et al. 2012; Wu 2005). In our study, through

481   comparing the two phylogenetic trees based on the conserved WRKY domains (Fig. 3) and

482   proteins (Fig. 4A), we obtained the nearly same classification of all SsWRKY genes, suggesting

483   that the conserved WRKY domain is an indispensable unit in WRKY genes. The variation of the

484   WRKYGQK heptapeptide may influence the proper DNA-binging ability of WRKY genes (Duan

485   et al. 2007; Maeo et al. 2001). A recent binding study by Brand et al. disclosed that a reciprocal

486   Q/K change of the WRKYGQK heptapeptide might result in different DNA-binding specificities

487 of the respective WRKY genes (Brand et al. 2013). For instance, the soybean WRKY genes,

488 GmWRKY6 and GmWRKY21, which contains the WRKYGKK variant, can't bind normally to

489 the W-box(Zhou et al. 2008). NtWRKY12 gene in tobacco with the WRKYGKK variant

490 recognizes another binding sequence 'TTTTCCAC' instead of normal W-box (van Verk et al.

491 2008). In our study, four WRKY genes (SsWRKY14, SsWRKY23, SsWRKY38 and

492 SsWRKY78) had a single mismatched amino acid in their conserved WRKYGQK heptapeptide

493 (Fig. 1). The variants detected in willow were extremely congruent with that in another

494 salicaceous plant, poplar, which also contains the same three variants in seven PtWRKY genes

495 (He et al. 2012). Previous studies have disclosed that the binding specificities of variable

496 WRKYGQK heptapeptide vary tremendously (Brand et al. 2013), however, few studies were

497 shown about the effect of variable zinc finger motif. In this study, four WRKY domains

498 (SsWRKY76C, SsWRKY64, SsWRKY12 and SsWRKY28) without complete zinc finger motif

499 may lack the ability of interacting with W-box, as well as PtWRKY83, 40, 95 and 10 in poplar

500 (He et al. 2012). Thereby, it is still indispensable to further investigate the function or the

501 expression patterns of the regulated gene targets in the variant sequences of the WRKY domains

502 (both WRKYGQK heptapeptide and complete zinc finger motif).

503   Different classification methods may lead to different numbers of WRKY genes in each group.

504 The classification method in our study was categorized as described in *Arabidopsis*, grape,

505 cucumber, castor bean and many other plant species (Eulgem 2000; Guo et al. 2014; Ling et al.

506 2011; Zou et al. 2016). According to this method, the willow WRKY genes were classified into

507 three main groups (I, II and III), with five subgroups in group II (IIa, IIb, IIc, IId and IIe).

508 However, the strategy described in rice and poplar was a little different (He et al. 2012; Wu

509 2005). They classified the subgroup IIc categorized above into a new subgroup Ib based on the

510 fact that the C-termini of group I and the domains of the above subgroup IIc shared more similar

511 consensus structures. At the meantime, subgroup IId and IIe categorized above were reclassified

512 into subgroup IIc and IId, respectively. With the same classification method as described in

513   *Arabidopsis* and many other plants, the numbers of different groups in poplar and rice are

514   illustrated in Table S5. WRKY genes of subgroup IIa, the smallest number of members, appear to

515   play crucial roles in regulating biotic and abiotic stress responses (Rushton et al. 2010). As shown

516   in Table S5, the willow WRKY genes of subgroup IIa and IIb are extremely similar to that of

517   other plant species, suggesting that all SsWRKY genes of these subgroups have been identified.

518   In addition, the numbers of WRKY III in eurosids I group, such as cucumber (6), poplar (10),

519   grape (6) and willow (7) are less than that of eurosids II (*Arabidopsis*: 14) and monocots (rice:

520   36), suggesting that different duplication events or selection pressures occurred in WRKY III

521   genes after the divergence of eurosids I and eurosids II group. A previous study in *Arabidopsis*

522   showed that nearly all WRKY III members respond to diverse biotic stresses, indicating that this

523   group probably evolved with the increasing biological requirements (Wang et al. 2015). The

524   different numbers of WRKY III genes in willow, poplar, cucumber, *Arabidopsis* and rice are

525   probably due to their different biotic stresses during evolution, and seven SsWRKY III genes

526   may be sufficient for the biological requirements in willow.

527   WRKY transcription factors play important roles in the regulation of developmental processes

528   and response to biotic and abiotic stress (Brand et al. 2013). The evolutionary relationship of

529   WRKY gene family promises to obtain significant insights into how biotic and abiotic stress

530   responses from single cellular aquatic algae to multicellular flowering plants (Rinerson et al.

531   2015). Previous studies hypothesized that group I WRKY genes were generated by domain

532   duplication of a proto-WRKY gene with a single WRKY domain, group II WRKY genes evolved

533   through the subsequent loss of N-terminal WRKY domain, and group III genes evolved from the

534   replacement of conserved His residue with a Cys residue in zinc motif (Wu 2005). However,

535   recent study proposed two alternative hypotheses of WRKY gene evolution (Rinerson et al.

536   2015): the "Group I Hypothesis" and the "IIa + b Separate Hypothesis". Additionally, another

537   recent study by Brand et al. concluded that subgroup IIc WRKY genes evolved directly from IIc-

538   like ancestral WRKY domains, and group I genes evolved independently due to a duplication of

539   the IIc-like ancestral WRKY domains (Brand et al. 2013). Phylogenetic analysis in our study

540   shows that subgroup IIc and group IC are evolutionarily close, as well as subgroups IIa and IIb,

541   subgroups IId and IIe, and this result is consistent with the conclusion drew by Brand et al.

542   (Brand et al. 2013). Additionally, the V-type introns of SsWRKY genes are only found in

543   subgroup IIa and IIb, while R-type introns are found in other groups except group IN. The results

544   are congruent with the "IIa + b Separate Hypothesis". Our results shown here provide important

545   reference for the further analyses on the accurate evolutionary relationship of WRKY gene

546   family.

547   Gene duplication events played prominent roles in a succession of genomic rearrangements and

548   expansions, and it is also the main motivation of plants evolution (Vision et al. 2000). The gene

549   family expansion occurs via three mechanisms: TDs, SDs and transposition events (Maher et al.

550   2006), and we only focused on the TDs and SDs in this study. In willow, a total of 66 SsWRKY

551   genes were identified to participate in gene duplication events, and all of these genes appeared to

552   have undergone SDs. Similarly, in poplar, only one homologous gene pair participated in TDs,

553   while 29 of 42 (69%) homologous gene pairs were determined to participate in SDs. The similar

554   WRKY gene expansion patterns in willow and poplar showed that SDs were the main factors in

555   the expansion of WRKY genes in woody plants. However, in cucumber, no gene duplication

556   events have occurred in CsWRKY gene evolution, probably because there were no recent whole-

557   genome duplication and tandem duplication in cucumber genome (Huang et al. 2009). In rice and

558   *Arabidopsis*, many WRKY genes were generated by TDs, which was incongruent with the

559   duplication events in willow, poplar and cucumber. The different WRKY gene expansion

560   patterns of the above plant species could be due to their different life habits and selection

561   pressures in a large scale, and it is still indispensable to be further investigated.

562   The WRKY gene family plays crucial roles in response to biotic and abiotic stresses, as well as

563   diverse physiological and developmental processes in plant species. Because of the lack of

564   researches on the function of willow WRKY genes, our study provided putative functions of

565   SsWRKY genes by comparing the orthologous genes between willow and *Arabidopsis*. The

566   details of the functions or regulations of AtWRKY genes can be obtained from TAIR

567   (http://www.arabidopsis.org/). For example, AtWRKY2, the ortholog to SsWRKY6, which

568   highly expressed in the five examined tissues, plays important roles in seed germination and post

569   germination growth (Jiang & Yu 2009). AtWRKY33, the ortholog to SsWRKY1, 35, 55 and 84,

570   influences the tolerance to NaCl, inc sensitivity to oxidative stress and abscisic acid (Jiang &

571   Deyholos 2009). A large number of AtWRKY genes, i.e. AtWRKY3, 4, 18, 53, 41, work in the

572   resistance to *Pseudomonas syringae* (Chen & Chen 2002; Higashi et al. 2008; Lai et al. 2008;

573   Murray et al. 2007), so do their orthologs in willow (SsWRKY42, 47, 39, 79, 20 and 70). Based

574   on the comparison of willow WRKY genes with their *Arabidopsis* orthologs, we could speculate

575   that the functional divergence of SsWRKY genes has played prominent roles in the responses to

576   various stresses.

577   **Conclusions**

578   Based on the recent released willow genome sequence and RNA-seq data, in this study, we

579   identified 85 SsWRKY proteins using bioinformatics approach. According to the phylogenetic

580   relationships and structural features of WRKY domains, all 85 SsWRKY genes were assigned to

581   the group I, group II (subgroup a-e) and group III. Three variations of the WRKYGQK

582   heptapeptide and the normal zinc finger motif in willow WRKY genes might execute some new

583   biological functions. Evolutionary analysis of SsWRKY III genes will be helpful for

584   understanding the evolution of WRKY III genes in plant. With the comparison of willow WRKY

585   genes with their *Arabidopsis* orthologs, breeding willow varieties with increased tolerance to

586   many adverse environments could be achieved using transgenic technology. Our results will be

587   not only beneficial to complete the functional and annotation information of WRKY genes family

588   in woody plants, but also provide interesting gene pools to be investigated for breeding and

589   genetic engineering purposes in woody plants.

Acknowledgments

591   The authors are deeply grateful to Prof. Tongming Yin, who provided the draft willow genome

592   and RNA-seq data from five tissues. The authors also thank Prof. Ning Ye, Prof. Qiaolin Ye and

593   Dr. Yiqing Xu for providing valuable suggestions and comments.

References

595   Bailey TL, Williams N, Misleh C, and Li WW. 2006. MEME: discovering and analyzing DNA
596   and protein sequence motifs. *Nucleic Acids Res* 34:W369-373. 10.1093/nar/gkl198
597   Bornberg-Bauer E. 1998. Computational approaches to identify leucine zippers. *Nucleic Acids*
598   *Res* 26:2740-2746. 10.1093/nar/26.11.2740
599   Brand LH, Fischer NM, Harter K, Kohlbacher O, and Wanke D. 2013. Elucidating the
600   evolutionary conserved DNA-binding specificities of WRKY transcription factors by
601   molecular dynamics and in vitro binding assays. *Nucleic Acids Res* 41:9764-9778.
602   10.1093/nar/gkt732
603   Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, and Madden TL. 2009.
604   BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. 10.1186/1471-
605   2105-10-421
606   Chen C, and Chen Z. 2002. Potentiation of developmentally regulated plant defense response by
607   AtWRKY18, a pathogen-induced Arabidopsis transcription factor. *Plant Physiol* 129:706-
608   716. 10.1104/pp.001057
609   Chen F, Mackey AJ, Vermunt JK, and Roos DS. 2007. Assessing performance of orthology
610   detection   strategies   applied   to   eukaryotic   genomes.   *PLoS   One*   2:e383.
611   10.1371/journal.pone.0000383
612   Cheng H, Li H, Deng Y, Xiao J, Li X, and Wang S. 2015. The WRKY45-2-WRKY13-WRKY42
613   Transcriptional Regulatory Cascade Is Required for Rice Resistance to Fungal Pathogen.
614   *Plant Physiol* 167:1087-1099.
615   Chothia C, Gough J, Vogel C, and Teichmann SA. 2003. Evolution of the protein repertoire.
616   *Science* 300:1701-1703. 10.1126/science.1085371

617   Ciolkowski I, Wanke D, Birkenbihl RP, and Somssich IE. 2008. Studies on DNA-binding
618   selectivity of WRKY transcription factors lend structural clues into WRKY-domain
619   function. *Plant Mol Biol* 68:81-92. 10.1007/s11103-008-9353-1

620   Cormack RS, Eulgem T, Rushton PJ, Köchner P, Hahlbrock K, and Somssich IE. 2002. Leucine
621   zipper-containing WRKY proteins widen the spectrum of immediate early elicitor-
622   induced WRKY transcription factors in parsley. *Biochimica et Biophysica Acta (BBA) -*
623   *Gene Structure and Expression* 1576:92-100. 10.1016/s0167-4781(02)00298-1

624   Dai X, Hu Q, Cai Q, Feng K, Ye N, Tuskan GA, Milne R, Chen Y, Wan Z, Wang Z, Luo W, Wang
625   K, Wan D, Wang M, Wang J, Liu J, and Yin T. 2014. The willow genome and divergent
626   evolution from poplar after the common genome duplication. *Cell Res* 24:1274-1277.
627   10.1038/cr.2014.83

628   Di-Qiu YU, Chen LG, and Zhang LP. 2006. Transcription Factor WRKY Superfamily:Origin,
629   Structure and Function. *Acta Botanica Yunnanica* 28:69-77.

630   Ding M, Chen J, Jiang Y, Lin L, Cao Y, Wang M, Zhang Y, Rong J, and Ye W. 2015. Genome-
631   wide investigation and transcriptome analysis of the WRKY gene family in Gossypium.
632   *Mol Genet Genomics* 290:151-171. 10.1007/s00438-014-0904-7

633   Dong J, Chen C, and Chen Z. 2003. Expression profiles of the ArabidopsisWRKY gene
634   superfamily during plant defense response. *Plant Mol Biol* 51:21-37.
635   10.1023/a:1020780022549

636   Du L, and Chen Z. 2008. Identification of genes encoding receptor-like protein kinases as
637   possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in
638   Arabidopsis. *The Plant Journal* 24:837-847. 10.1111/j.1365-313X.2000.00923.x

639   Duan MR, Nan J, Liang YH, Mao P, Lu L, Li L, Wei C, Lai L, Li Y, and Su XD. 2007. DNA
640   binding mechanism revealed by high resolution crystal structure of Arabidopsis thaliana
641   WRKY1 protein. *Nucleic Acids Res* 35:1145-1154. 10.1093/nar/gkm001

642   Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14:755-763.
643   10.1093/bioinformatics/14.9.755

644   Eulgem T. 2000. The WRKY superfamily of plant transcription factors. *Trends in Plant Science*
645   5:199-206. 10.1016/s1360-1385(00)01600-9

646   Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y,
647   Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M,
648   Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, and Zhang J. 2004.
649   Bioconductor: open software development for computational biology and bioinformatics.
650   *Genome Biol* 5:R80. 10.1186/gb-2004-5-10-r80

651   Giacomelli JI, Ribichich KF, Dezar CA, and Chan RL. 2010. Expression analyses indicate the
652   involvement of sunflower WRKY transcription factors in stress responses, and
653   phylogenetic reconstructions reveal the existence of a novel clade in the Asteraceae. *Plant*
654   *Science* 178:398-410. 10.1016/j.plantsci.2010.02.008

655   Gu ZL, Cavalcanti A, Chen FC, Bouman P, and Li WH. 2002. Extent of gene duplication in the
656   genomes of Drosophila, nematode, and yeast. *Molecular Biology and Evolution* 19:256-
657   262.
658   Guo C, Guo R, Xu X, Gao M, Li X, Song J, Zheng Y, and Wang X. 2014. Evolution and
659   expression analysis of the grape (Vitis vinifera L.) WRKY gene family. *J Exp Bot*
660   65:1513-1528. 10.1093/jxb/eru007
661   He GH, Xu JY, Wang YX, Liu JM, Li PS, Chen M, Ma YZ, and Xu ZS. 2016. Drought-
662   responsive WRKY transcription factor genes TaWRKY1 and TaWRKY33 from wheat
663   confer drought and/or heat resistance in Arabidopsis. *BMC Plant Biol* 16:1-16.
664   He H, Dong Q, Shao Y, Jiang H, Zhu S, Cheng B, and Xiang Y. 2012. Genome-wide survey and
665   characterization of the WRKY gene family in Populus trichocarpa. *Plant Cell Rep*
666   31:1199-1217. 10.1007/s00299-012-1241-0
667   Higashi K, Ishiga Y, Inagaki Y, Toyoda K, Shiraishi T, and Ichinose Y. 2008. Modulation of
668   defense signal transduction by flagellin-induced WRKY41 transcription factor in
669   Arabidopsis thaliana. *Mol Genet Genomics* 279:303-312. 10.1007/s00438-007-0315-0
670   Holub EB. 2001. The arms race is ancient history in Arabidopsis, the wildflower. *Nat Rev Genet*
671   2:516-527. 10.1038/35080508
672   Hu B, Jin J, Guo AY, Zhang H, Luo J, and Gao G. 2015. GSDS 2.0: an upgraded gene feature
673   visualization server. *Bioinformatics* 31:1296-1297. 10.1093/bioinformatics/btu817
674   Huang S, Gao Y, Liu J, Peng X, Niu X, Fei Z, Cao S, and Liu Y. 2012. Genome-wide analysis of
675   WRKY transcription factors in Solanum lycopersicum. *Mol Genet Genomics* 287:495-
676   513. 10.1007/s00438-012-0696-6
677   Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li
678   J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EA, Wu Y, Guo J, He J, Jia
679   Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan,
680   Wu Z, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK,
681   Kim JY, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li
682   M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li
683   J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y,
684   Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H,
685   Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, and Li S. 2009. The genome of the
686   cucumber, Cucumis sativus L. *Nat Genet* 41:1275-1281. 10.1038/ng.475
687   Ishiguro S, and Nakamura K. 1994. Characterization of a cDNA encoding a novel DNA-binding
688   protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding
689   for sporamin and ß-amylase from sweet potato. *MGG Molecular & General Genetics* 244.
690   10.1007/bf00282746
691   Jaffar MA, Song A, Faheem M, Chen S, Jiang J, Liu C, Fan Q, and Chen F. 2016. Involvement of
692   CmWRKY10 in Drought Tolerance of Chrysanthemum through the ABA-Signaling
693   Pathway. *International Journal of Molecular Sciences* 17.

694   Jang J-Y, Choi C-H, and Hwang D-J. 2010. The WRKY Superfamily of Rice Transcription
695   Factors. *The Plant Pathology Journal* 26:110-114. 10.5423/ppj.2010.26.2.110
696   Jiang M, Liu QE, Liu ZN, Li JZ, and He CM. 2016. Over-expression of a WRKY transcription
697   factor gene BoWRKY6 enhances resistance to downy mildew in transgenic broccoli
698   plants. *Australasian Plant Pathology*:1-8.
699   Jiang W, and Yu D. 2009. Arabidopsis WRKY2 transcription factor mediates seed germination
700   and postgermination arrest of development by abscisic acid. *BMC Plant Biol* 9:96.
701   10.1186/1471-2229-9-96
702   Jiang Y, and Deyholos MK. 2009. Functional characterization of Arabidopsis NaCl-inducible
703   WRKY25 and WRKY33 transcription factors in abiotic stresses. *Plant Mol Biol* 69:91-
704   105. 10.1007/s11103-008-9408-3
705   Johnson CS. 2002. TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat
706   Development Gene of Arabidopsis, Encodes a WRKY Transcription Factor. *The Plant
707   Cell Online* 14:1359-1375. 10.1105/tpc.001404
708   Kim CY, Vo KTX, Cong DN, Jeong DH, Lee SK, Kumar M, Kim SR, Park SH, Kim JK, and
709   Jeon JS. 2016. Functional analysis of a cold-responsive rice WRKY gene, OsWRKY71.
710   *Plant Biotechnology Reports* 10:13-23.
711   Kuzovkina YA, and Quigley MF. 2005. Willows Beyond Wetlands: Uses of Salix L. Species for
712   Environmental Projects. *Water Air & Soil Pollution* 162:183-204.
713   Lagace M, and Matton DP. 2004. Characterization of a WRKY transcription factor expressed in
714   late torpedo-stage embryos of Solanum chacoense. *Planta* 219:185-189. 10.1007/s00425-
715   004-1253-2
716   Lai Z, Vinod K, Zheng Z, Fan B, and Chen Z. 2008. Roles of Arabidopsis WRKY3 and WRKY4
717   transcription factors in plant responses to pathogens. *BMC Plant Biol* 8:68. 10.1186/1471-
718   2229-8-68
719   Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F,
720   Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, and Higgins DG. 2007. Clustal
721   W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.
722   10.1093/bioinformatics/btm404
723   Letunic I, Doerks T, and Bork P. 2015. SMART: recent updates, new developments and status in
724   2015. *Nucleic Acids Res* 43:D257-260. 10.1093/nar/gku949
725   Li H, and Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler
726   transform. *Bioinformatics* 25:1754-1760. 10.1093/bioinformatics/btp324
727   Li J, Brader G, Kariola T, and Palva ET. 2006. WRKY70 modulates the selection of signaling
728   pathways in plant defense. *Plant J* 46:477-491. 10.1111/j.1365-313X.2006.02712.x
729   Ling J, Jiang W, Zhang Y, Yu H, Mao Z, Gu X, Huang S, and Xie B. 2011. Genome-wide analysis
730   of WRKY gene family in Cucumis sativus. *BMC Genomics* 12:471. 10.1186/1471-2164-
731   12-471

732 Liu J, Chen X, Liang X, Zhou X, Yang F, Liu J, He SY, and Guo Z. 2016. Alternative splicing of
733   rice WRKY62 and WRKY76 transcription factor genes in pathogen defense. *Plant*
734   *Physiol*.
735 Liu JJ, and Ekramoddoullah AK. 2009. Identification and characterization of the WRKY
736   transcription factor family in Pinus monticola. *Genome* 52:77-88. 10.1139/G08-106
737 Maeo K, Hayashi S, Kojima-Suzuki H, Morikami A, and Nakamura K. 2001. Role of conserved
738   residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins.
739   *Biosci Biotechnol Biochem* 65:2428-2436. 10.1271/bbb.65.2428
740 Maher C, Stein L, and Ware D. 2006. Evolution of Arabidopsis microRNA families through
741   duplication events. *Genome Res* 16:510-519. 10.1101/gr.4680506
742 Mangelsen E, Kilian J, Berendzen KW, Kolukisaoglu UH, Harter K, Jansson C, and Wanke D.
743   2008. Phylogenetic and comparative gene expression analysis of barley (Hordeum
744   vulgare) WRKY transcription factor family reveals putatively retained functions between
745   monocots and dicots. *BMC Genomics* 9:194. 10.1186/1471-2164-9-194
746 McInerney EM, Rose DW, Flynn SE, Westin S, Mullen TM, Krones A, Inostroza J, Torchia J,
747   Nolte RT, Assa-Munt N, Milburn MV, Glass CK, and Rosenfeld MG. 1998. Determinants
748   of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation.
749   *Genes Dev* 12:3357-3368. 10.1101/gad.12.21.3357
750 Murray SL, Ingle RA, Petersen LN, and Denby KJ. 2007. Basal resistance against Pseudomonas
751   syringae in Arabidopsis involves WRKY53 and a protein with homology to a nematode
752   resistance protein. *Mol Plant Microbe Interact* 20:1431-1438. 10.1094/MPMI-20-11-1431
753 Muthamilarasan M, Bonthala VS, Khandelwal R, Jaishankar J, Shweta S, Nawaz K, and Prasad
754   M. 2015. Global analysis of WRKY transcription factor superfamily in Setaria identifies
755   potential candidates involved in abiotic stress signaling. *Front Plant Sci* 6:910.
756   10.3389/fpls.2015.00910
757 Ohno S, Wolf U, and Atkin NB. 2009. Evolution from Fish to Mammals by Gene Duplication.
758   *Hereditas* 59:169-187. 10.1111/j.1601-5223.1968.tb02169.x
759 Overbeek R, Fonstein M, D'Souza M, Pusch GD, and Maltsev N. 1999. The use of gene clusters
760   to infer functional coupling. *Proceedings of the National Academy of Sciences* 96:2896-
761   2901. 10.1073/pnas.96.6.2896
762 Pleguezuelo CRR, Zuazo VHD, Bielders C, Bocanegra JAJ, Pereatorres F, and Martínez JRF.
763   2014. Bioenergy farming using woody crops. A review. *Agronomy for Sustainable*
764   *Development* 35:95-119.
765 Pnueli L, Hallak-Herr E, Rozenberg M, Cohen M, Goloubinoff P, Kaplan A, and Mittler R. 2002.
766   Molecular and biochemical mechanisms associated with dormancy and drought tolerance
767   in the desert legume Retama raetam. *The Plant Journal* 31:319-330. 10.1046/j.1365-
768   313X.2002.01364.x
769 Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G,
770   Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, and Finn RD.

771   2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290-301.
772   10.1093/nar/gkr1065
773 Ramamoorthy R, Jiang SY, Kumar N, Venkatesh PN, and Ramachandran S. 2008. A
774   comprehensive transcriptional profiling of the WRKY gene family in rice under various
775   abiotic and phytohormone treatments. *Plant Cell Physiol* 49:865-879.
776   10.1093/pcp/pcn061
777 Rinerson CI, Rabara RC, Tripathi P, Shen QJ, and Rushton PJ. 2015. The evolution of WRKY
778   transcription factors. *BMC Plant Biol* 15:66. 10.1186/s12870-015-0456-y
779 Robatzek S, and Somssich IE. 2002. Targets of AtWRKY6 regulation during plant senescence
780   and pathogen defense. *Genes Dev* 16:1139-1149. 10.1101/gad.222702
781 Ross CA, Liu Y, and Shen QJ. 2007. The WRKY Gene Family in Rice (Oryza sativa). *Journal of*
782   *Integrative Plant Biology* 49:827-842. 10.1111/j.1744-7909.2007.00504.x
783 Rossberg M, Theres K, Acarkan A, Herrero R, Schmitt T, Schumacher K, Schmitz G, and
784   Schmidt R. 2001. Comparative Sequence Analysis Reveals Extensive Microcolinearity in
785   the Lateral Suppressor Regions of the Tomato, Arabidopsis, and Capsella Genomes. *Plant*
786   *Cell* 13:979. 10.2307/3871354
787 Rushton PJ, Somssich IE, Ringler P, and Shen QJ. 2010. WRKY transcription factors. *Trends*
788   *Plant Sci* 15:247-258. 10.1016/j.tplants.2010.02.006
789 Sun C. 2003. A Novel WRKY Transcription Factor, SUSIBA2, Participates in Sugar Signaling in
790   Barley by Binding to the Sugar-Responsive Elements of the iso1 Promoter. *The Plant Cell*
791   *Online* 15:2076-2092. 10.1105/tpc.014597
792 Suyama M, Torrents D, and Bork P. 2006. PAL2NAL: robust conversion of protein sequence
793   alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609-612.
794   10.1093/nar/gkl315
795 Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: Molecular
796   Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30:2725-
797   2729. 10.1093/molbev/mst197
798 Ulker B, and Somssich IE. 2004. WRKY transcription factors: from DNA binding towards
799   biological function. *Curr Opin Plant Biol* 7:491-498. 10.1016/j.pbi.2004.07.012
800 van Verk MC, Pappaioannou D, Neeleman L, Bol JF, and Linthorst HJ. 2008. A Novel WRKY
801   transcription factor is required for induction of PR-1a gene expression by salicylic acid
802   and bacterial elicitors. *Plant Physiol* 146:1983-1995. 10.1104/pp.107.112789
803 Vision TJ, Brown DG, and Tanksley SD. 2000. The Origins of Genomic Duplications in
804   Arabidopsis. *Science* 290:2114-2117. 10.1126/science.290.5499.2114
805 Wagner GP, Kin K, and Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq
806   data: RPKM measure is inconsistent among samples. *Theory Biosci* 131:281-285.
807   10.1007/s12064-012-0162-3
808 Wang Y, Feng L, Zhu Y, Li Y, Yan H, and Xiang Y. 2015. Comparative genomic analysis of the
809   WRKY III gene family in populus, grape, arabidopsis and rice. *Biology direct* 10:1.

810   Wei W, Zhang Y, Han L, Guan Z, and Chai T. 2008. A novel WRKY transcriptional factor from
811   Thlaspi caerulescens negatively regulates the osmotic stress tolerance of transgenic
812   tobacco. *Plant Cell Rep* 27:795-803.
813   Wu KL. 2005. The WRKY Family of Transcription Factors in Rice and Arabidopsis and Their
814   Origins. *DNA Research* 12:9-26. 10.1093/dnares/12.1.9
815   Xie Z, Zhang ZL, Zou X, Huang J, Ruas P, Thompson D, and Shen QJ. 2005. Annotations and
816   functional analyses of the rice WRKY gene superfamily reveal positive and negative
817   regulators of abscisic acid signaling in aleurone cells. *Plant Physiol* 137:176-189.
818   10.1104/pp.104.054312
819   Xiong W, Xu X, Zhang L, Wu P, Chen Y, Li M, Jiang H, and Wu G. 2013. Genome-wide analysis
820   of the WRKY gene family in physic nut (Jatropha curcas L.). *Gene* 524:124-132.
821   10.1016/j.gene.2013.04.047
822   Xu X, Chen C, Fan B, and Chen Z. 2006. Physical and functional interactions between pathogen-
823   induced Arabidopsis WRKY18, WRKY40, and WRKY60 transcription factors. *Plant*
824   *Cell* 18:1310-1326. 10.1105/tpc.105.037523
825   Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T, Yabuki T, Aoki M, Seki E, Matsuda T,
826   Tomo Y, Hayami N, Terada T, Shirouzu M, Tanaka A, Seki M, Shinozaki K, and
827   Yokoyama S. 2005. Solution structure of an Arabidopsis WRKY DNA binding domain.
828   *Plant Cell* 17:944-956. 10.1105/tpc.104.026435
829   Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and*
830   *Evolution* 24:1586-1591. 10.1093/molbev/msm088
831   Zhang J. 2003. Evolution by gene duplication: an update. *Trends in Ecology & Evolution* 18:292-
832   298. 10.1016/s0169-5347(03)00033-8
833   Zhang Y, and Wang L. 2005. The WRKY transcription factor superfamily: its origin in eukaryotes
834   and expansion in plants. *BMC Evol Biol* 5:1. 10.1186/1471-2148-5-1
835   Zhang ZL, Xie Z, Zou X, Casaretto J, Ho TH, and Shen QJ. 2004. A rice WRKY gene encodes a
836   transcriptional repressor of the gibberellin signaling pathway in aleurone cells. *Plant*
837   *Physiol* 134:1500-1513. 10.1104/pp.103.034967
838   Zhou QY, Tian AG, Zou HF, Xie ZM, Lei G, Huang J, Wang CM, Wang HW, Zhang JS, and
839   Chen SY. 2008. Soybean WRKY-type transcription factor genes, GmWRKY13,
840   GmWRKY21, and GmWRKY54, confer differential tolerance to abiotic stresses in
841   transgenic Arabidopsis plants. *Plant Biotechnol J* 6:486-503. 10.1111/j.1467-
842   7652.2008.00336.x
843   Zou X, Seemann JR, Neuman D, and Shen QJ. 2004. A WRKY gene from creosote bush encodes
844   an activator of the abscisic acid signaling pathway. *J Biol Chem* 279:55770-55779.
845   10.1074/jbc.M408536200
846   Zou Z, Yang L, Wang D, Huang Q, Mo Y, and Xie G. 2016. Gene Structures, Evolution and
847   Transcriptional Profiling of the WRKY Gene Family in Castor Bean (Ricinus communis
848   L.). *PLoS One* 11:e0148243. 10.1371/journal.pone.0148243

Figure 1

Comparison of the WRKY domain sequences from 85 SsWRKY genes.

The WRKY gene with the suffix -N and -C indicates the N-terminal and C-terminal WRKY domain of group I members, respectively. "-" has been inserted for the optimal alignment. Red indicates the highly conserved WRKYGQK heptapeptide, and the zinc finger motifs are highlighted in green. The position of a conserved intron is indicated by an arrowhead.

Group I N

Group I C

Group II a

Group II b

Group II c

Group II d

Group II e

Group III

Figure 2

Chromosomal location of SsWRKY genes.

Red indicates group I, blue indicates group II and green indicates group III. The chromosome numbers are given at the top of each chromosome and the left side of each chromosome is related to the approximate physical location of each WRKY gene. Only one unmapped SsWRKY gene is shown on ChrN.

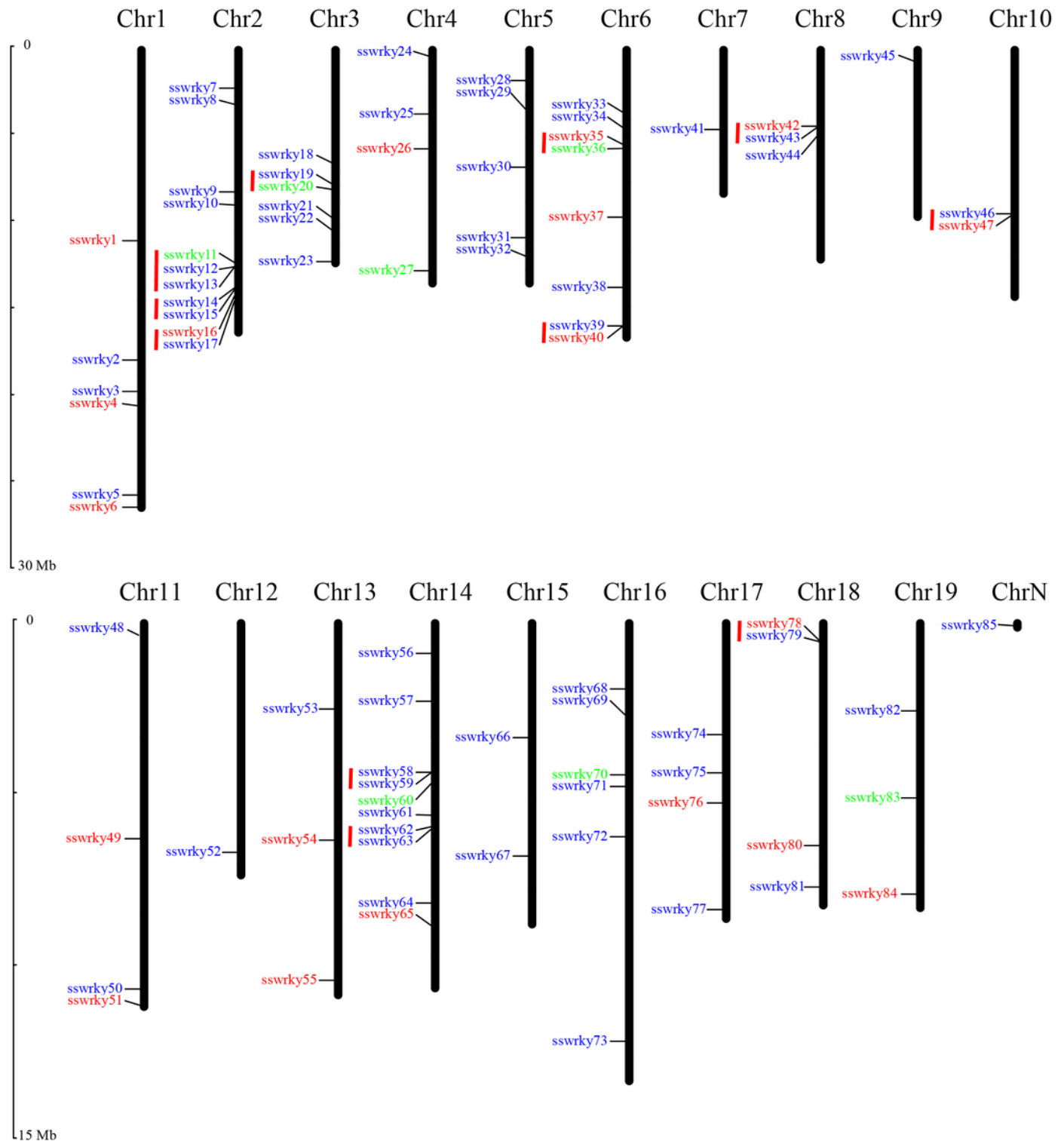



Figure 3

Phylogenetic tree of WRKY domains from willow and *Arabidopsis*.

he phylogenetic tree was constructed using the neighbor-joining method in MEGA 6.0. The WRKY genes with the suffix 'N' and 'C' indicate the N-terminal and the C-terminal WRKY domains of group I, respectively. The different colors indicate different groups (I, II and III) or subgroups (IIa, b, c, d and e) of WRKY domains. Circles indicate WRKY genes from willow, and diamonds represent genes from *Arabidopsis*. The purple trapezoid region indicate a new subgroup belonging to IIc.

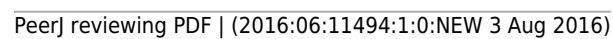



Figure 4

Genomic organization of SsWRKY genes.

A) The phylogenetic tree built on the basis of full-length SsWRKY genes was depicted using the neighbor-joining method in MEGA 6.0. The short black lines indicate the existence of duplicated gene pairs; B) The graphic exon-intron structure of SsWRKY genes is displayed using GSDS. Green indicates exons, and gray indicates introns. The introns phases 0, 1 and 2 are indicated by numbers 0, 1 and 2, respectively.

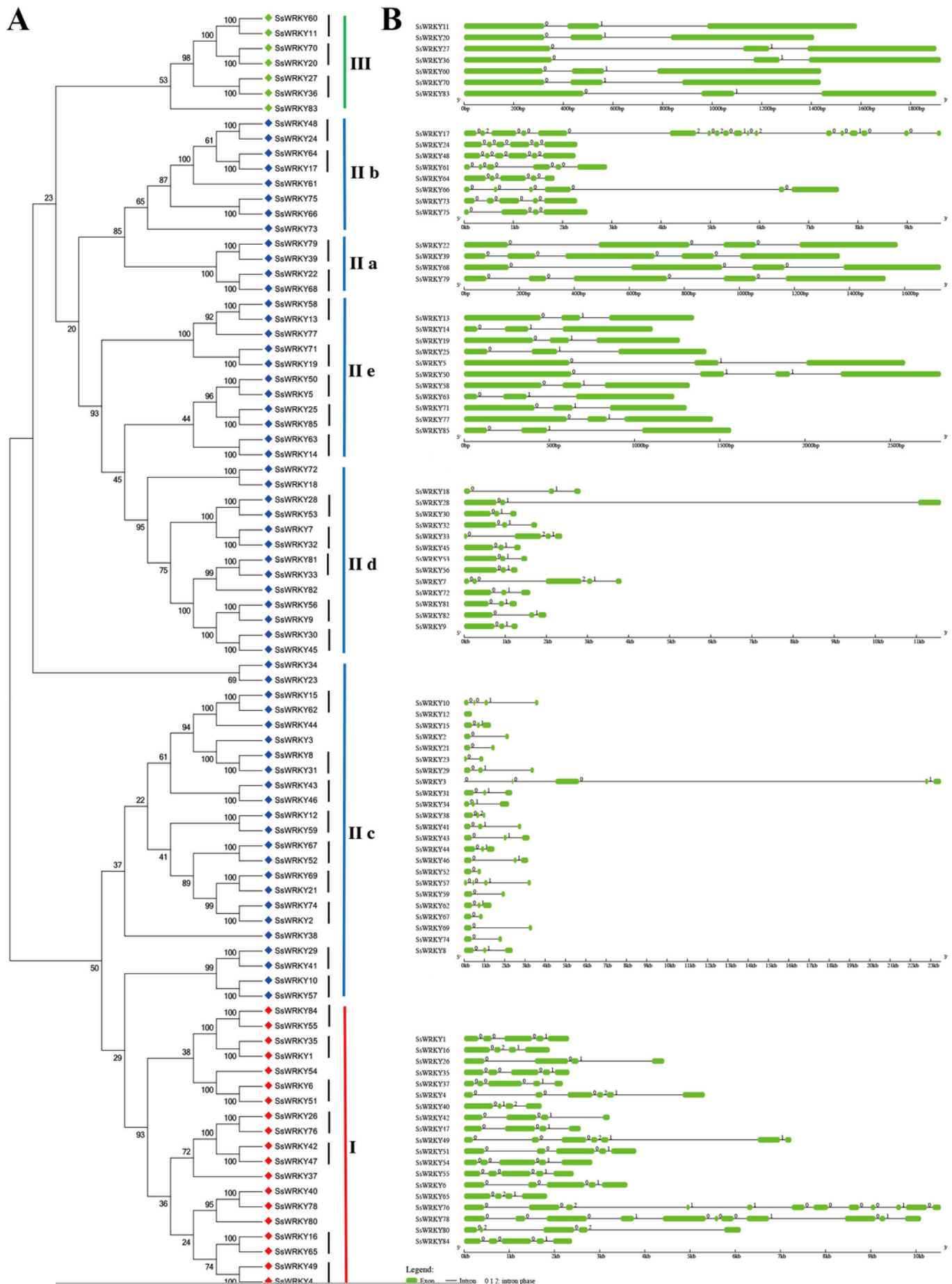
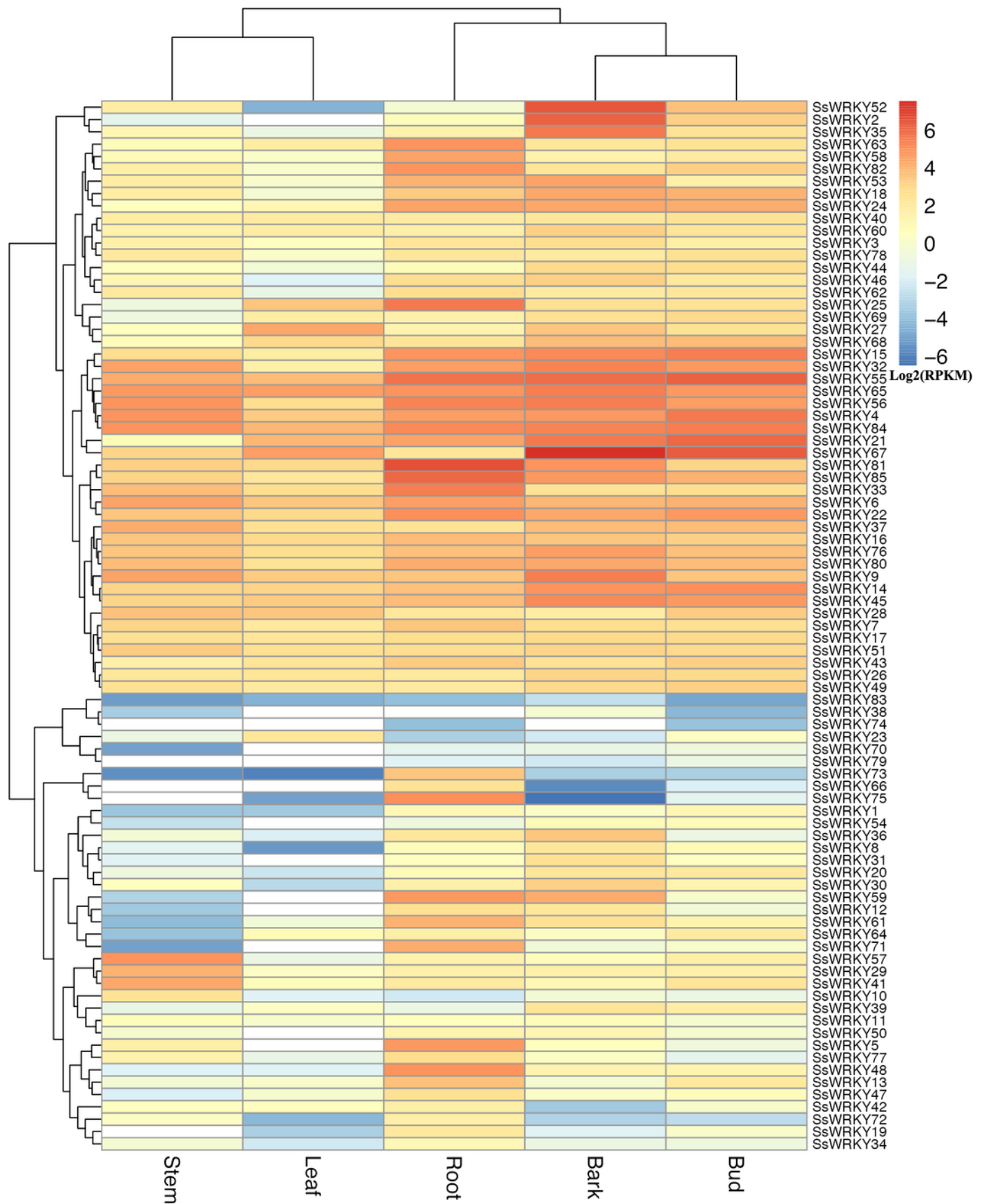



Figure 5

Expression profiles of the 85 SsWRKY genes in root, stem, bark, bud and leaf.

Color scale represents RPKM normalized log2 transformed counts and red indicates high expression, blue indicates low expression and white indicates the gene is not expressed in this tissue.

The detailed characteristics of WRKY genes identified in willow.

1   **Table 1. The detailed characteristics of WRKY genes identified in willow.**

Gene	SequenceID	Chr	Group	Ortholog		Deduced polypeptide			Introns
				AtWRKY	PtWRKY	Length(aa)	PI	MW(kDa)	
SsWRKY1	willow_GLEAN_10011238	1	I	33	17	583	7.14	64.7	4
SsWRKY2	willow_GLEAN_10019192	1	II c	45	43	162	9.47	18.6	1
SsWRKY3	willow_GLEAN_10017208	1	II c	28,71	29	584	9.42	65.6	4
SsWRKY4	willow_GLEAN_10017139	1	I	20	44	560	6.99	60.9	5
SsWRKY5	willow_GLEAN_10007860	1	II e	35	45	445	5.92	48.4	2
SsWRKY6	willow_GLEAN_10003806	1	I	2	37,101,102	733	5.69	78.8	4
SsWRKY7	willow_GLEAN_10022392	2	II d	21	46,63	453	9.53	49.9	4
SsWRKY8	willow_GLEAN_10022273	2	II c	71	47	328	6.89	37.0	2
SsWRKY9	willow_GLEAN_10009329	2	II d	15	14,94	339	9.77	37.5	2
SsWRKY10	willow_GLEAN_10009231	2	II c	12	48	204	7.64	23.6	3
SsWRKY11	willow_GLEAN_10016913	2	III	30	6,51	351	6.27	39.2	2
SsWRKY12	willow_GLEAN_10016886	2	II c	-	19,50	129	6.75	14.6	0
SsWRKY13	willow_GLEAN_10016883	2	II e	22	23,49,78	352	5.81	38.3	2
SsWRKY14	willow_GLEAN_10019911	2	II e	-	3	247	5.58	28.1	2
SsWRKY15	willow_GLEAN_10019925	2	II c	23	13,33	319	6.46	35.6	2
SsWRKY16	willow_GLEAN_10019982	2	I	1	54	472	6.88	52.2	3
SsWRKY17	willow_GLEAN_10020022	2	II b	47	53	1081	5.25	116.8	17
SsWRKY18	willow_GLEAN_10025583	3	II d	-	55	142	9.60	16.5	2
SsWRKY19	willow_GLEAN_10025423	3	II e	29	41	335	5.54	37.9	2
SsWRKY20	willow_GLEAN_10025378	3	III	41/53	21	342	5.25	38.4	2
SsWRKY21	willow_GLEAN_10008020	3	II c	45	18	157	9.41	17.8	1
SsWRKY22	willow_GLEAN_10006448	3	II a	40	88	320	8.38	35.4	3
SsWRKY23	willow_GLEAN_10013342	3	II c	-	39	109	8.03	12.9	1
SsWRKY24	willow_GLEAN_10009960	4	II b	42	28,79	604	6.93	65.3	5
SsWRKY25	willow_GLEAN_10017267	4	II e	65	8,58	267	5.43	29.7	2
SsWRKY26	willow_GLEAN_10018559	4	I	58	60	537	8.72	58.9	3
SsWRKY27	willow_GLEAN_10004854	4	III	54	85	323	5.70	36.3	2
SsWRKY28	willow_GLEAN_10008312	5	II d	-	-	490	10.27	54.0	2
SsWRKY29	willow_GLEAN_10009112	5	II c	13	68	235	8.70	26.7	2
SsWRKY30	willow_GLEAN_10003565	5	II d	15	20	310	9.48	34.3	2
SsWRKY31	willow_GLEAN_10016009	5	II c	28,71	62	322	6.67	36.2	2
SsWRKY32	willow_GLEAN_10018195	5	II d	21	46,63	349	9.69	38.8	2
SsWRKY33	willow_GLEAN_10026833	6	II d	7	91	339	9.89	36.8	3
SsWRKY34	willow_GLEAN_10026721	6	II c	49	34	287	5.25	32.1	2
SsWRKY35	willow_GLEAN_10026591	6	I	33	64	572	6.41	62.7	4
SsWRKY36	willow_GLEAN_10026566	6	III	54	85	329	6.13	36.7	2

SsWRKY37	willow_GLEAN_10020588	6	I	44	93	478	9.25	52.5	4
SsWRKY38	willow_GLEAN_10026166	6	II c	51	67	233	5.03	26.1	2
SsWRKY39	willow_GLEAN_10026455	6	II a	18/60	9	327	9.02	36.2	4
SsWRKY40	willow_GLEAN_10026458	6	I	32	15	413	8.26	44.9	3
SsWRKY41	willow_GLEAN_10008192	7	II c	13	68	236	9.21	26.6	2
SsWRKY42	willow_GLEAN_10025108	8	I	3/4	69	460	8.80	50.6	3
SsWRKY43	willow_GLEAN_10025123	8	II c	57	71	295	6.32	32.3	2
SsWRKY44	willow_GLEAN_10015641	8	II c	48	70	357	6.11	39.9	2
SsWRKY45	willow_GLEAN_10008155	9	II d	15	20,26	331	9.57	36.4	2
SsWRKY46	willow_GLEAN_10013562	10	II c	57	71	289	6.26	31.9	2
SsWRKY47	willow_GLEAN_10013586	10	I	3/4	72	490	8.60	53.7	3
SsWRKY48	willow_GLEAN_10004012	11	II b	42	100	585	6.48	63.3	5
SsWRKY49	willow_GLEAN_10006060	11	I	20	44	607	7.09	6.6	6
SsWRKY50	willow_GLEAN_10007614	11	II e	35	74	481	5.39	51.6	3
SsWRKY51	willow_GLEAN_10007542	11	I	2	37	734	6.10	79.7	4
SsWRKY52	willow_GLEAN_10013801	12	II c	-	75	178	9.08	20.5	1
SsWRKY53	willow_GLEAN_10012158	13	II d	74	25	356	9.66	40.0	2
SsWRKY54	willow_GLEAN_10004417	13	I	2	35	697	6.52	76.1	4
SsWRKY55	willow_GLEAN_10007732	13	I	33	1	602	7.65	66.0	4
SsWRKY56	willow_GLEAN_10009039	14	II d	15	14,94	362	9.39	40.0	2
SsWRKY57	willow_GLEAN_10016668	14	II c	12	48	180	8.47	20.7	3
SsWRKY58	willow_GLEAN_10016177	14	II e	22	23,49,78	354	6.35	38.8	2
SsWRKY59	willow_GLEAN_10016180	14	II c	43	19,50	193	9.47	21.7	1
SsWRKY60	willow_GLEAN_10016220	14	III	30	6	368	5.03	41.3	2
SsWRKY61	willow_GLEAN_10018940	14	II b	42	28,79	467	8.78	50.0	5
SsWRKY62	willow_GLEAN_10018891	14	II c	23	13,33	318	5.71	35.6	2
SsWRKY63	willow_GLEAN_10018881	14	II e	-	80	263	5.05	29.7	2
SsWRKY64	willow_GLEAN_10020302	14	II b	36	-	460	6.28	50.0	4
SsWRKY65	willow_GLEAN_10020380	14	I	1	2	481	5.98	52.8	3
SsWRKY66	willow_GLEAN_10011119	15	II b	9	99	618	6.55	66.2	5
SsWRKY67	willow_GLEAN_10016438	15	II c	-	82	178	9.35	20.5	1
SsWRKY68	willow_GLEAN_10023347	16	II a	40	88	320	8.82	35.3	3
SsWRKY69	willow_GLEAN_10023447	16	II c	45	18	178	9.17	20.1	1
SsWRKY70	willow_GLEAN_10023687	16	III	41/53	21	336	5.17	37.2	2
SsWRKY71	willow_GLEAN_10023735	16	II e	29	41	325	5.54	36.6	2
SsWRKY72	willow_GLEAN_10014752	16	II d	-	55	338	9.24	37.9	2
SsWRKY73	willow_GLEAN_10009602	16	II b	9	42	509	5.51	55.3	4
SsWRKY74	willow_GLEAN_10010473	17	II c	45	43	182	9.92	20.9	1
SsWRKY75	willow_GLEAN_10015128	17	II b	9	86	544	6.01	59.0	3

SsWRKY76	willow_GLEAN_10015184	17	I	58	87	1044	8.94	116.1	11
SsWRKY77	willow_GLEAN_10005468	17	II e	27	96	411	5.96	45.7	2
SsWRKY78	willow_GLEAN_10006860	18	I	-	90	1593	8.67	179.0	10
SsWRKY79	willow_GLEAN_10006862	18	II a	18/60	9	320	8.57	35.6	4
SsWRKY80	willow_GLEAN_10011608	18	I	32	-	528	5.74	57.8	4
SsWRKY81	willow_GLEAN_10004546	18	II d	7	7,91	300	9.80	32.8	2
SsWRKY82	willow_GLEAN_10003422	19	II d	11/17	24	339	9.58	37.1	2
SsWRKY83	willow_GLEAN_10011321	19	III	55	36,76	358	5.63	38.7	2
SsWRKY84	willow_GLEAN_10005288	19	I	33	4	597	6.69	65.6	4
SsWRKY85	willow_GLEAN_10002834	N/A	II e	65	58	268	5.83	30.2	2

2   Chr, chromosome numbers.

3   N/A, not available.

4   "-", not detected.