

Genome-wide Identification of *WRKY* Genes in the Desert Poplar *Populus euphratica* and Adaptive Evolution of the Genes in Response to Salt Stress

Jianchao Ma, Jing Lu, Jianmei Xu, Bingbing Duan, Xiaodong He and Jianquan Liu

State Key Laboratory of Grassland Agro-Ecosystem, School of Life Science, Lanzhou University, Lanzhou, Gansu, People's Republic of China.

Supplementary Issue: RNA: An Expanding View of Function and Evolution

ABSTRACT: *WRKY* transcription factors play important roles in plant development and responses to various stresses in plants. However, little is known about the evolution of the *WRKY* genes in the desert poplar species *Populus euphratica*, which is highly tolerant of salt stress. In this study, we identified 107 *PeWRKY* genes from the *P. euphratica* genome and examined their evolutionary relationships with the *WRKY* genes of the salt-sensitive congener *Populus trichocarpa*. Ten *PeWRKY* genes are specific to *P. euphratica*, and five of these showed altered expression under salt stress. Furthermore, we found that two pairs of orthologs between the two species showed evidence of positive evolution, with dN/dS ratios >1 (nonsynonymous/synonymous substitutions), and both of them altered their expression in response to salinity stress. These findings suggested that both the development of new genes and positive evolution in some orthologs of the *WRKY* gene family may have played an important role in the acquisition of high salt tolerance by *P. euphratica*.

KEYWORDS: *WRKY* genes, expansion, *Populus euphratica*, adaptive evolution

SUPPLEMENT: RNA: An Expanding View of Function and Evolution

CITATION: Ma et al. Genome-wide Identification of *WRKY* Genes in the Desert Poplar *Populus euphratica* and Adaptive Evolution of the Genes in Response to Salt Stress. *Evolutionary Bioinformatics* 2015;11(S1) 47–55 doi: 10.4137/EBO.S22067.

TYPE: Review

RECEIVED: May 12, 2015. **RESUBMITTED:** July 01, 2015. **ACCEPTED FOR PUBLICATION:** July 03, 2015.

ACADEMIC EDITOR: Jike Cui, Associate Editor

PEER REVIEW: Three peer reviewers contributed to the peer review report. Reviewers' reports totaled 1,179 words, excluding any confidential comments to the academic editor.

FUNDING: The research was supported by the National Key Project for Basic Research (2012CB114504) and the National High-Tech Research and Development Program of China (863 Program, No. 2013AA102605). The authors confirm that the funder had no influence over the study design, content of the article, or selection of this journal.

COMPETING INTERESTS: Authors disclose no potential conflicts of interest.

CORRESPONDENCE: liujq@lzu.edu.cn

COPYRIGHT: © the authors, publisher and licensee Libertas Academica Limited. This is an open-access article distributed under the terms of the Creative Commons CC-BY-NC 3.0 License.

Paper subject to independent expert blind peer review. All editorial decisions made by independent academic editor. Upon submission manuscript was subject to anti-plagiarism scanning. Prior to publication all authors have given signed confirmation of agreement to article publication and compliance with all applicable ethical and legal requirements, including the accuracy of author and contributor information, disclosure of competing interests and funding sources, compliance with ethical requirements relating to human and animal study participants, and compliance with any copyright requirements of third parties. This journal is a member of the Committee on Publication Ethics (COPE).

Published by Libertas Academica. Learn more about this journal.

Introduction

Transcription factors (TFs) regulating gene expression via a modular structure containing at least one DNA-binding domain play pivotal roles in signaling and regulatory networks.^{1,2} The *WRKY* TF family in plants comprises numerous members that regulate genes involved in seed germination, seed dormancy, trichome development, lignin biosynthesis, and both biotic and abiotic stress responses.^{3–8} Since the first *WRKY* gene was cloned and characterized from sweet potato, numerous *WRKY* genes have been identified in other plants, including *Arabidopsis*, rice, cucumber, tomato, grape, and poplar.^{9–15} Proteins encoded by the *WRKY* gene family contain two prominent and defining structural characteristics, *WRKY* domains and zinc-finger motifs, and *WRKY* proteins are divided into three groups (I–III) on the basis of the numbers and types of these motifs that they contain.^{16–18} Group I contains two *WRKY* domains and $CX_4CX_{22-23}HXH$ zinc-finger motifs, while the Group II members have the same type of finger motif as Group I, C_2H_2 (in this case $C-X_{4-5}-C-X_{22-23}-H-X_1-H$) and contain one *WRKY* domain; the *WRKY* domain of the third group (III) comprises a C_2HC ($C-X_7-C-X_{23}-H-X_1-C$) motif.¹⁰ Mutations in either the *WRKY* domain or the zinc finger of a *WRKY* domain will result in the

loss of DNA-binding ability.¹⁶ *WRKY* proteins bind to the W-box ((C/T)TGAC(T/C)) elements of their target genes^{17,18} as well as to a *cis*-element (SURE).¹⁹

Some *WRKY* genes have been reported to be activated in response to various abiotic stresses including high salinity.^{8,18} Salinity is a major abiotic stress, acting primarily by changing osmotic stress and disrupting cellular homeostasis and ion distribution. It also induces oxidative stress, damages membranes and proteins, and activates signaling cascades regulating gene expression.²⁰ The regulatory networks that underlie salt tolerance are very complex.²¹ To date, a small number of *WRKY* genes have been identified as taking part in this process in plants. For instance, it has been reported that *OsWRKY45* plays a part in rice salt tolerance,²² while *GhWRKY39* has a positive role in cotton salt tolerance.²³ Overexpression of a maize *WRKY58* gene enhances salt tolerance in transgenic rice.²⁴ Similarly, in *Arabidopsis*, overexpression of either *AtWRKY25* or *AtWRKY33* increases salt tolerance.²⁵ A recent study also showed that some *PeWRKY* genes are involved in salt tolerance.²⁶ However, few studies have been designed in order to examine changes in the number of *WRKY* genes and adaptive evolution of orthologous genes between species that are closely related but show contrasting degrees of salt tolerances.



In this study, we aimed to examine these issues using *Populus euphratica* and *Populus trichocarpa*. The model tree species *P. trichocarpa* grows in mesophytic habitats, while *P. euphratica* grows in deserts; it is well adapted to salt stress and has a high level of salt tolerance.^{27–29} Genome sequences for both species are available^{30,31} and preliminary analyses suggested that the expression of some *WRKY* genes is altered in response to salinity stress.³⁰ The *WRKY* genes in the *P. trichocarpa* genome have been identified.¹⁵ In order to better understand the adaptive evolution of the *WRKY* genes in these two poplars of contrasting salt tolerance, we identified the *WRKY* genes of *P. euphratica* and examined the expression under salt stress of both species-specific genes and those that are orthologous in the two species. Our results provide further information about the roles that the *WRKY* genes may have played in the adaptive history of the desert poplar.

Methods

Identification of *WRKY* genes in two poplar genomes.

Candidate *WRKY* proteins in the *P. euphratica* genome³⁰ were identified using the BlastP program (P value = $1e-3$). Sequences of *Arabidopsis* members of the *WRKY* protein family downloaded from the *Arabidopsis* genome website TAIR 9.0 (<http://www.Arabidopsis.org/index.jsp>) were used as query sequences. In addition, the Hidden Markov Model³² profile for the *WRKY* domain (PF03106.7) from the Pfam database³³ was applied as a query against the *P. euphratica* genome using the program HMMER³⁴ in order to identify *WRKY* sequences by searching for the *WRKY*-binding domain. To verify the authenticity of these candidate sequences, the Pfam database was used to confirm that each candidate *PeWRKY* protein was a member of the *WRKY* family. The gene model IDs of *P. trichocarpa* *WRKY* genes were obtained from the Plant Transcription Factor Database v3.0 (<http://planttfdb.cbi.pku.edu.cn/index.php>) and confirmed using the same method as above. The *PeWRKYs* and *PtWRKYs* were named according to their position on the scaffolds and chromosomes of the respective genomes.

Mapping the *PeWRKY* genes onto the *P. trichocarpa* chromosomes. To determine the chromosomal locations of potential poplar *WRKY* genes, the *WRKY* orthologs in *P. trichocarpa* were used as references. Data on gene position, gene length, chromosome size, and position relative to the centromere of *PtWRKYs* were obtained from the Phytozome v9.1 website (<http://www.phytozome.net/poplar>). The MScanX software package³⁵ was used to complete the chromosomal distribution of *PeWRKY* genes. The map of the distribution of *PeWRKYs* on 19 chromosomes was drawn, based on the above data, using custom-designed Perl scripts.

Phylogenetic analyses of *WRKY* genes. Alignments of *WRKY* protein and domain sequences were carried out separately using MUSCLE³⁶ and then adjusted manually before the construction of phylogenetic trees. Phylogenetic trees were constructed based on the aligned *WRKY*-binding

domains and full predicted protein sequences of the *PeWRKY* genes using CLUSTAL³⁷ and MEGA 6,³⁸ respectively. The neighbor-joining (NJ) method was used with the following parameters: Poisson correction, pairwise deletion, and bootstrap (1,000 replicates). Exons and introns were predicted by comparing the coding sequences with the corresponding genomic sequences using custom-designed Perl scripts. The MEME program³⁹ was used to predict conserved motifs, with the following parameters: maximum number of motifs, 20; minimum motif width, 6; and maximum motif width, 70.

Divergence of orthologs between *P. euphratica* and *P. trichocarpa*. Identification of genes orthologous between *P. euphratica* and *P. trichocarpa* was performed using the bidirectional best hit method, as described in detail by Zhang et al.^{40,41} The MScanX software toolkit and the phylogenetic tree were also used to provide further confirmation. Orthologous genes were aligned by the Probabilistic Alignment Kit software package.⁴² To detect divergence of orthologs between *P. euphratica* and *P. trichocarpa*, a maximum likelihood method in PAML was applied to calculate nonsynonymous (K_a) and synonymous (K_s) substitution ratios using the YN00 program.⁴³ The elements in the promoter fragments (from -1500 to $+1$ bp) were predicted by PlantCARE online (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>).

Gene expression analysis. We collected Illumina RNA-Seq data using the methods described previously by Ma et al.³⁰ Calli induced from *P. euphratica* and *P. trichocarpa* shoots were collected from Tarim Basin desert in Xinjiang and the United States, respectively. RNA-Seq transcriptomic data from *P. euphratica* and *P. trichocarpa* calli under salt stress treatments (200 mM NaCl for 0, 6, 12, 24, and 48 hours) were generated using the Illumina HiSeqTM2000 platform. Three biological replicates were carried out. Gene expression levels were calculated as reads per kilobase of exon model per million mapped reads (RPKM).⁴⁴ RPKM values were \log_2 -transformed and used for heat map generation using the pheatmap package in the R language.

Results and Discussion

Identification of the *WRKY* genes in *P. euphratica*. The entire *P. euphratica* genome was searched by BlastP analysis using 70 *Arabidopsis* *WRKY* genes as query sequences. A total of 107 full-length putative *WRKY* genes were identified. To confirm that these were *WRKY* genes, the Pfam database was used to search for *WRKY* domains in each; as a result, all 107 (that were named *PeWRKY01* to *PeWRKY107*) were confirmed as *WRKY* genes (Table S1). The conserved *WRKY* domain, which contains approximately 60 amino acid residues, is considered to be a crucial element that usually interacts with the W-box (C/T)TGAC(T/C) to activate a large number of defense-related genes.^{15,18} Mutations in the *WRKYGQK* motif will result in the loss of DNA-binding ability.¹⁶ Most *WRKY* genes contain one *WRKY* domain except 23 Group I genes, which contain two *WRKY* domains. Multiple sequence

alignments of the 128 WRKY domains showed that 119 contained the conserved amino acid sequence WRKYGQK, while the others had one or two mismatched amino acids in this domain (Fig. S1): WRKYGHK (*PeWRKY38*), WRKYGRK (*PeWRKY71_N*), WAKYGQK (*PeWRKY94*), WRKDGQK (*PeWRKY40_C*), WKKYGQK (*PeWRKY17*), WRKYGKK (*PeWRKY102, 12, 28*), and WRK- (*PeWRKY81*).

Chromosomal distribution of *PeWRKY* genes on the reference poplar genome. *PeWRKYs* were widely distributed within the *P. trichocarpa* reference genome (Fig. 1). We found that 97 of the 107 *PeWRKY* genes could be mapped to 18 of the 19 *P. trichocarpa* chromosomes (Chr) according to the positions of orthologous *PtWRKYs* or linked genes (Table S2). Thirteen *WRKY* genes were found on Chr 1, 12 on Chr 2, and 11 on Chr 14. In contrast, Chrs 12, 15, and 19 each possess only two of the genes. We recovered 10 tandemly duplicated gene pairs that contained 20 *PeWRKY* genes on these chromosomes. We also found that 9 of these 10 duplicated gene pairs were shared with *P. trichocarpa* (Fig S2), suggesting that they existed before the divergence of the two species. We found one tandemly duplicated gene pair (*PeWRKY06* and *PeWRKY07*) specific to *P. euphratica* and one pair (*PtWRKY31* and *PtWRKY32*) specific to *P. trichocarpa*; these probably either formed after the divergence of the two poplars or represent ancestral sequences retained in one species but lost in the other.

Exon–intron organization and motif identification in *PeWRKY* genes. We next examined the exon–intron structure and motifs of all the *PeWRKY* genes identified (Fig. 2). Most *PeWRKY* genes have two to four introns: 56 out of 107 *PeWRKYs* possess two introns, 14 have three introns, and 15 have four introns. However, 8 *PeWRKY* genes have only

one intron and 13 genes have five introns, while only 1 has six introns. Changes in exon–intron organization may have contributed to evolutionary divergence between the different members of the gene family.⁴⁵

In addition, the online Multiple Expectation Maximization for Motif Elicitation program was used to find conserved motifs in the *PeWRKY* proteins (Table S4). As illustrated in Figure 2, other than motifs 1, 2, 3, and 5, which characterize the WRKY domains, *PeWRKY* members of the same subgroup were generally found to share a common motif composition. For instance, motif 17 is specific to 8 out of 12 genes in Group III while motif 10 and motif 15 are only present in Subgroup II d. Both Subgroup II a and Subgroup II b contain motif 7; Subgroup II c could be divided into two sections, of which one contains motif 14, which is specific to Subgroup II c and the other only contains motifs 1, 2, and 19, which comprise WRKY domains. Members of Subgroup II d mostly shared a conserved CaM-binding domain (motif 10), which is associated with calmodulin in *Arabidopsis*,⁴⁶ implying that these WRKY proteins may play pivotal roles in signal transduction by regulating the activity of target proteins. In general, the motifs within each group were relatively consistent and are probably associated with the similar functions carried out by members of that subgroup.

Phylogenetic analyses. To examine the evolutionary relationships among the WRKY domains, phylogenetic relationships were constructed using the amino acid sequences of the conserved domains of *AtWRKYs*, *PtWRKYs*, and *PeWRKYs* (Fig. 3) and the protein sequences encoded by *PeWRKYs* (Fig. 2). Based on *AtWRKY* classification and *PeWRKY* sequences, all the *PeWRKY* genes identified could be divided into three tentative groups. The first group (Group I) comprised

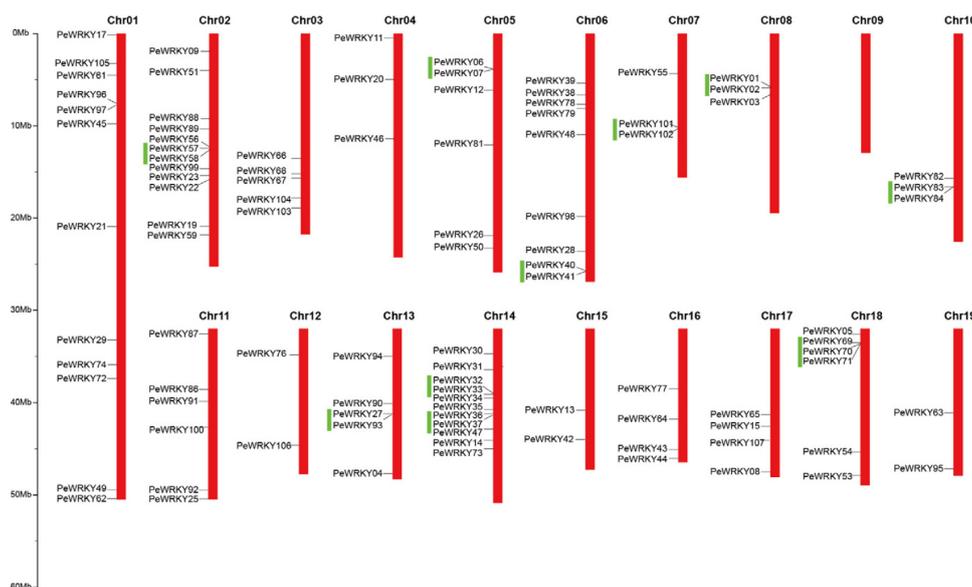


Figure 1. Chromosomal distributions of *PeWRKY* genes in the *Populus trichocarpa* reference genome and chromosomes. The chromosome number is indicated at the top of each chromosome. Tandemly duplicated genes are indicated with vertical green lines. Gene position and the size of each chromosome can be estimated using the scale on the left of the figure.

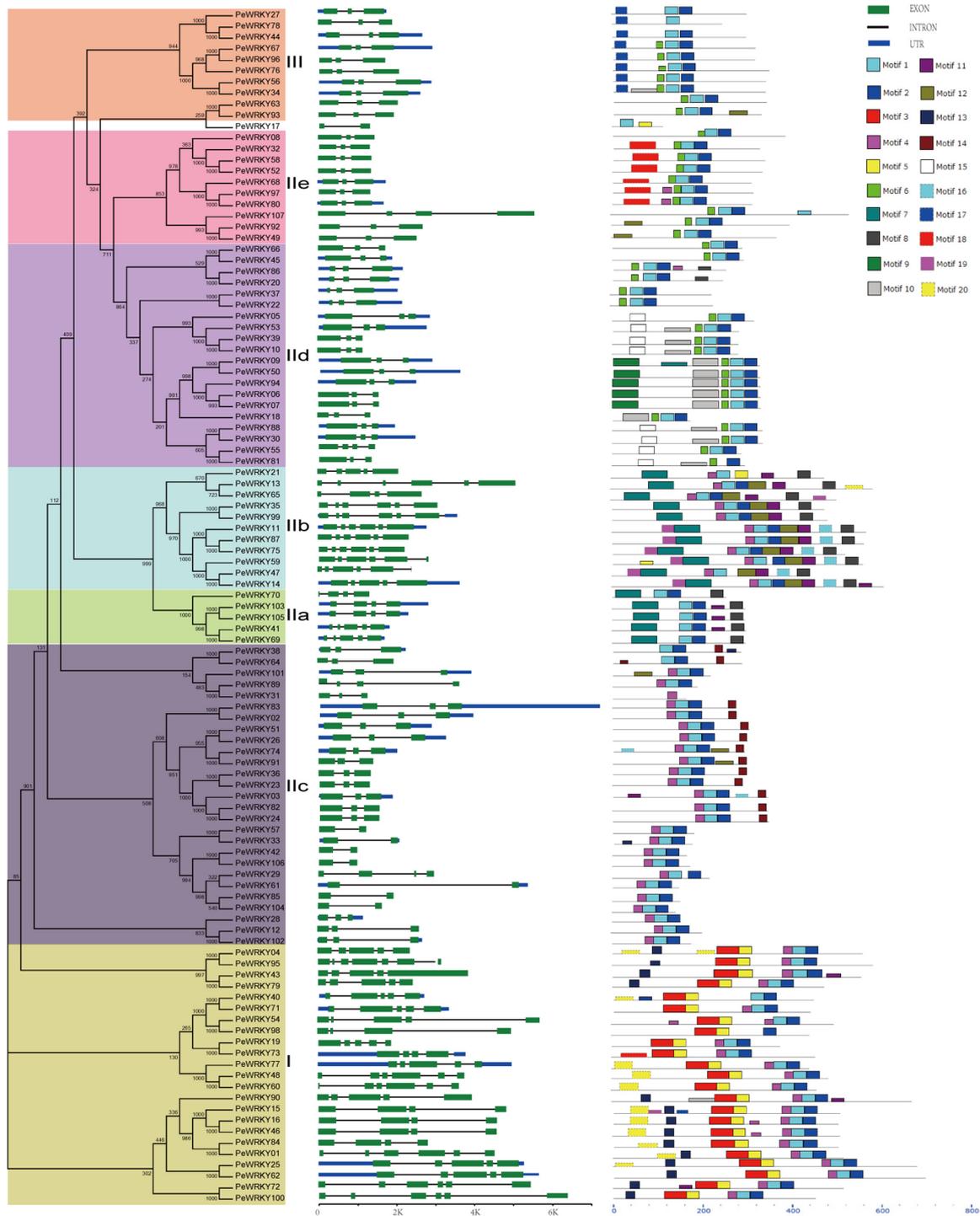


Figure 2. Phylogenetic relationship, intron–exon structures, and conserved motif compositions in the *PeWRKY* family. The unrooted phylogenetic tree was aligned by MUSCLE and constructed using MEGA 6 by the maximum likelihood method with 1,000 bootstrap replicates. The tree was divided into seven phylogenetic subgroups, indicated with different colored backgrounds. The intron–exon structures of 107 *PeWRKY* genes are shown in the middle panel. The right panel is a schematic representation of the conserved motifs in the *PeWRKY* proteins as detected by MEME analysis, represented by different colored boxes. Narrow rectangles representing the *E* values of motif are not significant as others. The length of each protein can be estimated using the scale at the bottom.

23 members, most of which contained two WRKY domains. Group II, which had one WRKY domain could be divided into five subgroups (Subgroups IIa–e), which contained 5, 11, 27, 16, and 14 members, respectively. The third group contained the remaining 10 members, which contained the WRKY domains

and C₂HC zinc-finger structure (C–X₄–C–X₂₃–H–X₁–C).⁴⁷ One gene (*PeWRKY17*), which lacked this structure, clustered weakly with the third group. The three groups recovered here were consistent with the previous classification of *WRKYs* in *Arabidopsis*¹⁰ and grape.⁴⁸ The ortholog of *PeWRKY17* in

P. trichocarpa was found to have lost the characteristic structure of Group III WRKYs.⁴⁹ The small group of IIC genes in Figure 3 (orthologs of *AtWRKY49*) that separated from the larger group were mainly the result of domain sequences and also agreed with the result in grape.⁴⁸ In addition, two genes (*PeWRK71* and *PeWRKY98*) of Group I were found to have lost the WRKYGQK motif at the C-terminal, according to the WRKY domains predicted by the Pfam database (Fig. S1). Because genes of Group I have been found to exist in all plants and in some organisms with ancient origins, they may be ancestral to the other two groups and subgroups.⁴⁷

Through phylogenetic analysis, we found 10 *PeWRKY* genes specific to *P. euphratica* while *P. trichocarpa* has five species-specific genes. One *P. euphratica*-specific gene was formed by tandem duplication, while the other nine genes in this species, although located in different scaffolds, show high sequence similarity to the other *PeWRKY* genes and cluster together with them in the phylogenetic tree. The latter nine genes probably either originated by segmental duplication after the two species diverged or were retained in *P. euphratica* but lost in *P. trichocarpa* after genome duplication in the ancestor of the two species.

Comparison of the number of WRKY genes in the different groups in other species with those in *P. euphratica* shows that among the *PeWRKY*s, duplication occurred mainly in Group I and Subgroups IIb and IId. We also found five unique genes (*PtWRKY*s) specific to *P. trichocarpa*. These results indicated that either the WRKY gene family expanded separately in the two species or some members were selectively retained in each popular species after genome duplication in their common ancestor.

Evolutionary divergence of orthologs between *P. euphratica* and *P. trichocarpa*. To examine the divergence of orthologs between *PeWRKY*s and *PtWRKY*s, nonsynonymous substitutions (dN) and synonymous substitutions (dS) per site between orthologous genes were calculated: dN/dS = 1, <1, and >1 indicate, respectively, neutral selection, purifying selection, and accelerated evolution with positive selection.⁵⁰ A total of 97 pairs of orthologous genes were identified, and the dN/dS value of each was determined (Table S1). The results indicated that most of the gene pairs (95/97) showed dN/dS ratios <1, showing that purifying selection predominated (Table S3) as is the case for other gene families. However, we found that two pairs of genes (*PeWRKY45* ver-

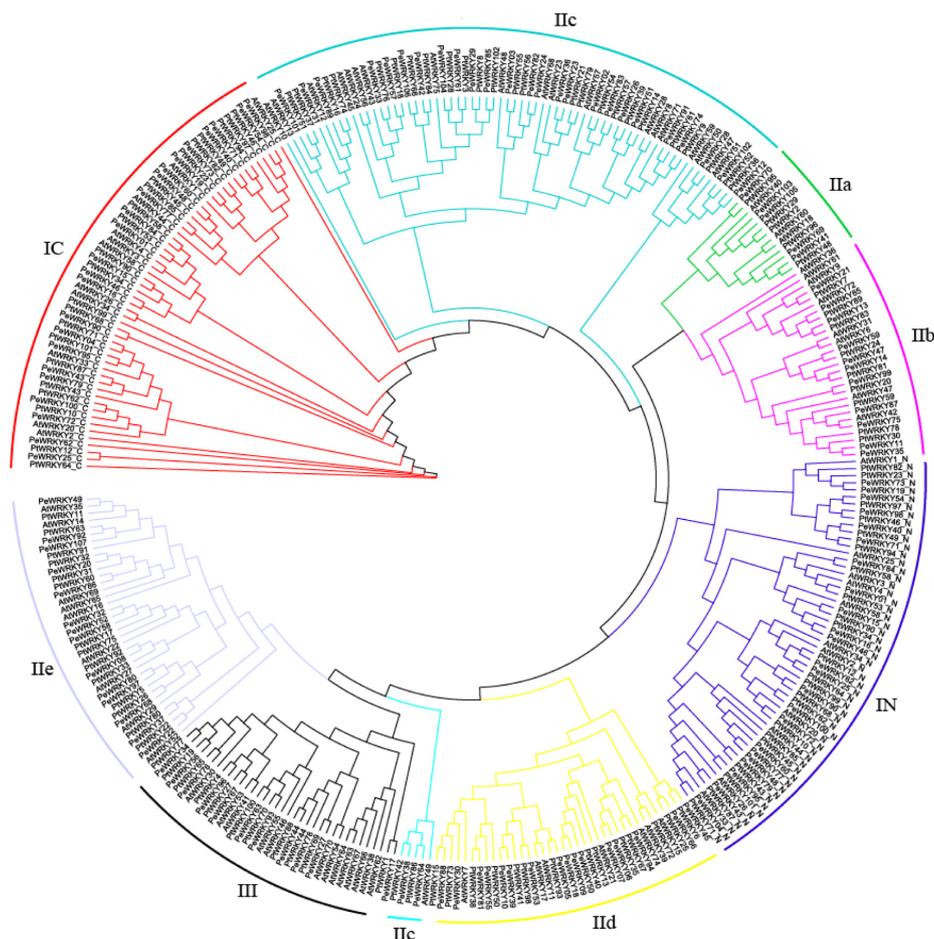


Figure 3. Phylogenetic tree of WRKY domains of *Arabidopsis*, *Populus trichocarpa*, and *Populus euphratica*. Multiple sequence alignment of WRKY domains was done using MUSCLE, and the phylogenetic tree was constructed using CLUSTAL with 1,000 bootstrap replicates by the NJ method. Group I proteins with the suffix N or C indicate N-terminal WRKY domains or C-terminal WRKY domains, respectively. The tree was divided into eight phylogenetic subgroups, designated as IC, IN, IIa–e, and III. Groups and subgroups are shown in different colors.



versus *PtWRKY6* and *PeWRKY102* versus *PtWRKY52*) had dN/dS ratios >1, suggesting that positive evolution was likely for these two genes.

Expression data showed that the positive evolved gene pairs showed different expression levels under salt stress in two poplar species (Fig. S3, Fig. 4). We examined the functions of the *Arabidopsis* orthologs of *PeWRKY45* and *PeWRKY102* and

AtWRKY17 and *AtWRKY50*. Both of them negatively regulate jasmonic acid (JA) signaling.^{51,52} JA concentration have also been found to be associated with salt tolerance in *P. euphratica*.⁵³ We further found that the promoter regions of these pairs of genes varied slightly (Table S5–S6). Taken together, these results suggest that *PeWRKY45* and *PeWRKY102* probably mediate the response to salt stress by increasing the level

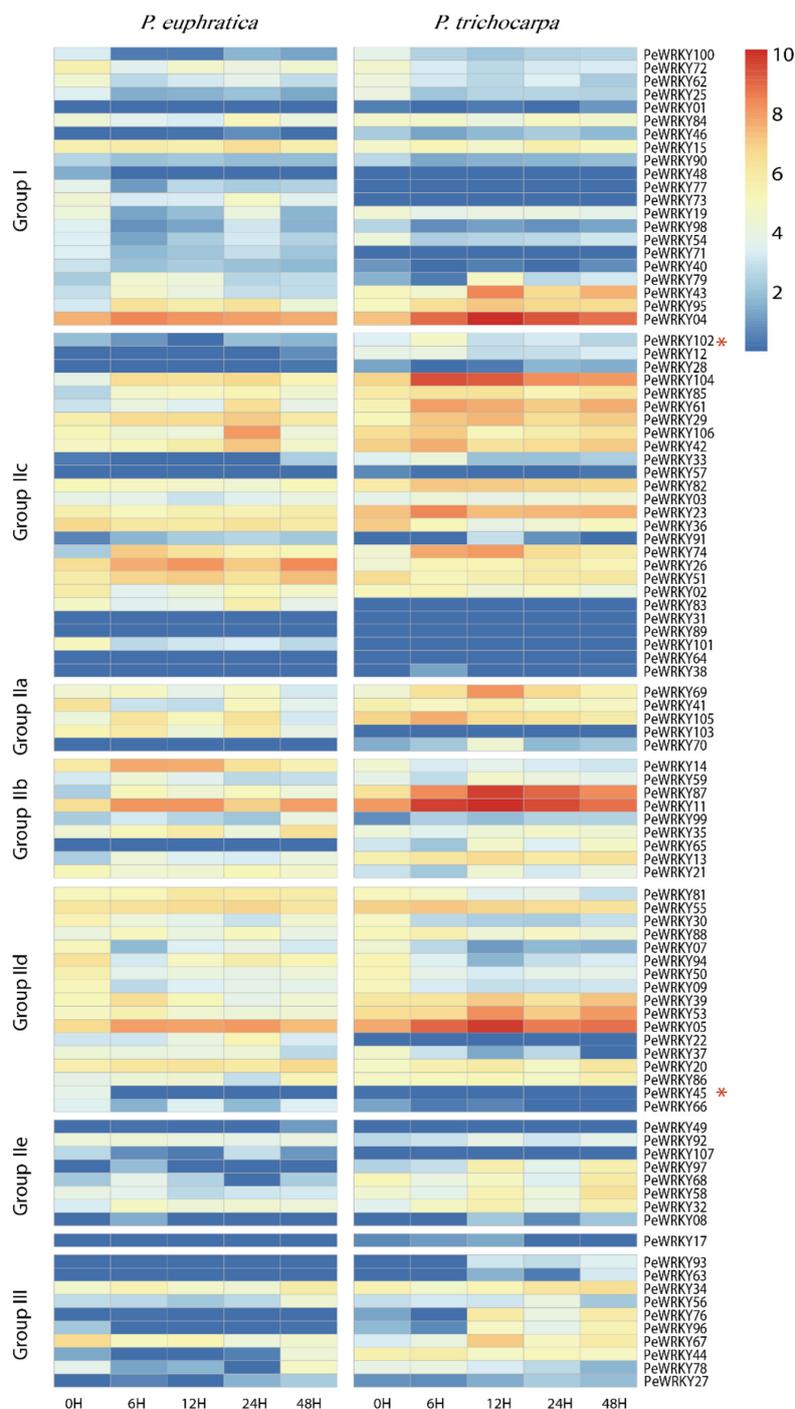


Figure 4. Heat map representation of expression of orthologous genes between *PeWRKYs* and *PtWRKYs* under salt stress. The Illumina RNA-Seq data were analyzed, and the RPKM values were log₂-transformed and a heat map generated. The left and right panels represent RNA-Seq data from *Populus euphratica* and *Populus trichocarpa*, respectively. Blue and red colors are used to represent low and high levels of expression, respectively. Adaptive evolved genes were marked by red asterisk.

Table 1. Size of the WRKY groups and subgroups in different plant species.

GENE	GROUP I	GROUP II					GROUP III	NG	TOTAL
		IIA	IIB	IIC	IID	IIE			
<i>AtWRKY</i> ⁵	15	3	8	17	7	8	14	0	72
<i>OsWRKY</i> ⁶	15	4	8	15	7	11	36	0	96
<i>VvWRKY</i> ⁴⁸	12	3	8	15	6	7	6	2	59
<i>PtWRKY</i>	22	5	9	27	14	14	10	1	102
<i>PeWRKY</i>	23	5	11	27	16	14	10	1	107

of expression of JA-related genes, and this may shed light on how *P. euphratica* tolerates salt stress.

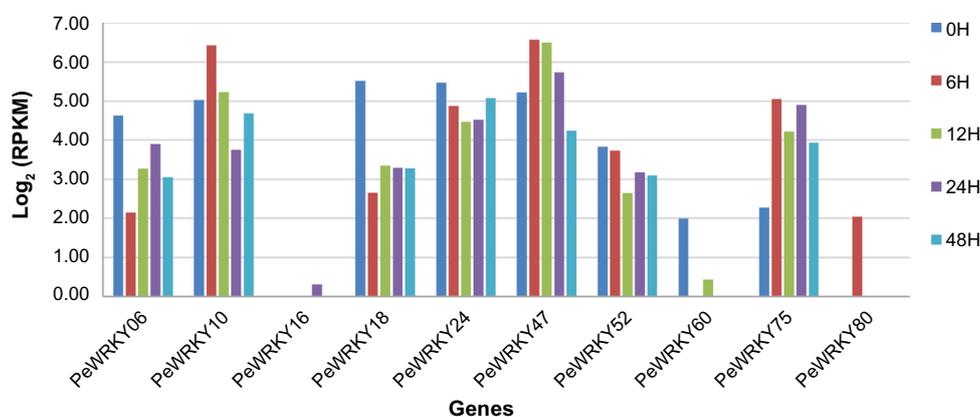
Expression profiling of PeWRKYs and PtWRKYs under salt stress. A total of 92/107 *PeWRKY* genes and 84/102 *PtWRKY* genes were detected as being expressed in calli with or without salt stress (Table S7). Of 92 *PeWRKY* genes, 23 showed changes in expression in response to salt stress: most (16/23) were downregulated while seven were upregulated. However, only five of the *PtWRKY* genes altered expression in response to salt stress. Further analysis showed that a total of 11 *PeWRKY* genes, among which are included two pairs of orthologs (*PeWRKY45* versus *PtWRKY6* and *PeWRKY102* versus *PtWRKY52*) that show evidence of adaptive evolution (Fig. S3), underwent different changes in expressions in response to salt stress compared to their *P. trichocarpa* orthologs (Fig. 4). Of the 23 *PeWRKY* genes differentially expressed under salt stress, nine belong to Group I and five to Subgroup IID while the others are members of the remaining groups or subgroups. Members of Subgroup IID mostly share the conserved CaM-binding domain that is associated with calmodulin in *Arabidopsis*.⁴⁶ The upregulated genes include the one reported to mediate salt stress in *P. euphratica*.²⁶ These findings suggest that members of Group I and Subgroup IID may play more roles in response to salt stress than members of the other groups of genes.

It is believed that Group III of the WRKY gene family evolved more recently than the other two groups.⁴⁷ Many members of this group in *Arabidopsis*, including *AtWRKY30*, *53*, *54*, and *70*, have been identified as regulators of senescence.⁵⁴ They are involved in a number of signaling pathways in response to different stresses. We calculated the dN/dS ratios for *WRKY* genes in the two *Populus* species; the average ratio in Group III was found to be 0.42 while the average in the other groups was 0.37, a finding consistent with the above assertion. In addition, no difference in expression was detected for Group III members when *P. trichocarpa* was subjected to salt stress.¹⁵ However, in the present study we found that *PeWRKY78*, a member of this group, was downregulated, whereas its *P. trichocarpa* ortholog *PtWRKY44* was slightly upregulated under salt stress.

Of the 10 *PeWRKY* genes unique to *P. euphratica*, three were downregulated and two were upregulated in response to salt stress (Fig. 5); this supports the notion that the origin or retention of these species-specific genes may have been driven by the need for high salt tolerance in this desert poplar. Overall, the results suggest that *WRKY* genes play an important role in salt stress tolerance in *P. euphratica*.

Conclusion

A total of 107 *WRKY* genes were identified in *P. euphratica* with 10 of them being species-specific genes that had originated through tandem and segmental duplication. When members of the WRKY family were compared between *P. euphratica* and *P. trichocarpa*, two pairs of genes exhibited evidence of positive selection (dN/dS > 1); they are probably significant in the adaptation of this poplar species to salt stress. Expression profiling of the *PeWRKY* genes revealed that five genes specific to *P. euphratica* and two that showed evidence of adaptive evolution showed alterations in expression in response to salt stress. These findings suggested that the formation of new genes and the occurrence of adaptive evolution in other genes may have played an important role in the evolution of high salt tolerance in *P. euphratica*.


Figure 5. Expression levels of 10 *PeWRKY* genes specific to *Populus euphratica* under salt stress. The y-axis shows the log₂-transformed RPKM values of genes expressed in callus after different periods of salt stress, and the x-axis shows the duplicated *WRKY* genes.



Author Contributions

Conceived and designed the experiments: JLi. Analyzed the data: JM. Wrote the first draft of the manuscript: JM. Contributed to the writing of the manuscript: JLu, JX, XH. Agree with manuscript results and conclusions: JLi, JM. Jointly developed the structure and arguments for the paper: JLi, XH. Made critical revisions and approved final version: JM, XH, BD. All authors reviewed and approved of the final manuscript.

Supplementary Material

Supplementary Table S1. Identification of *WRKY* TF genes in *Populus euphratica*.

Supplementary Table S2. *WRKY* TF genes in *Populus trichocarpa*.

Supplementary Table S3. Divergent selection between *PeWRKY* and *PtWRKY* genes.

Supplementary Table S4. The details of the motif sequences.

Supplementary Tables S5–S6. *Cis*-acting elements analysis of positive evolved genes.

Supplementary Table S7. RNA-seq data of *PeWRKYs* and *PtWRKYs* in calli.

Supplementary Figure S1. Multiple sequence alignment of *WRKY* domains using DNAMAN.

Supplementary Figure S2. Chromosome distribution of *PeWRKYs* and *PtWRKYs*.

Supplementary Figure S3. Expression levels of positive evolutionary orthologous genes.

REFERENCES

- Tombuloglu H, Kekec G, Sakcali MS, Unver T. Transcriptome-wide identification of R2R3-MYB transcription factors in barley with their boron responsive expression analysis. *Mol Genet Genomics*. 2013;288:141–55.
- Jang J-Y, Choi CH, Hwang D-J. The WRKY superfamily of rice transcription factors. *Plant Pathol J*. 2010;26:110–14.
- Wenck AR, Conger BV, Trigiano RN, Sams CE. Inhibition of somatic embryogenesis in orchardgrass by endogenous cytokinins. *Plant Physiol*. 1988;88:990–2.
- Pnueli L, Hallak-Herr E, Rozenberg M, et al. Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume *Retama raetam*. *Plant J*. 2002;31:319–30.
- Johnson CS, Kolevski B, Smyth DR. TRANSPARENT TESTA GLABRA 2, a trichome and seed coat development gene of *Arabidopsis*, encodes a WRKY transcription factor. *Plant Cell*. 2002;14:1359–75.
- Guillaumie S, Mzid R, Méchin V, et al. The grapevine transcription factor *WRKY2* influences the lignin pathway and xylem development in tobacco. *Plant Mol Biol*. 2010;72:215–34.
- Wang H, Avci U, Nakashima J, Hahn MG, Chen F, Dixon RA. Mutation of WRKY transcription factors initiates pith secondary wall formation and increases stem biomass in dicotyledonous plants. *Proc Natl Acad Sci USA*. 2010;107:22338–43.
- Eulgem T, Somssich IE. Networks of WRKY transcription factors in defense signaling. *Curr Opin Plant Biol*. 2007;10:366–71.
- Ishiguro S, Nakamura K. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and β -amylase from sweet potato. *Mol Gen Genet*. 1994;244:563–71.
- Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. *Trends Plant Sci*. 2000;5:199–206.
- Xie Z, Zhang ZL, Zou X, et al. Annotations and functional analyses of the rice *WRKY* gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol*. 2005;137:176–89.
- Ling J, Jiang W, Zhang Y, et al. Genome-wide analysis of WRKY gene family in *Cucumis sativus*. *BMC Genomics*. 2011;12:471.
- Huang S, Gao Y, Liu J, et al. Genome-wide analysis of WRKY transcription factors in *Solanum lycopersicum*. *Mol Genet Genomics*. 2012;287:495–513.
- Guo C, Guo R, Xu X, et al. Evolution and expression analysis of the grape (*Vitis vinifera* L.) *WRKY* gene family. *J Exp Bot*. 2014;65:1513–28.
- He H, Dong Q, Shao Y, et al. Genome-wide survey and characterization of the *WRKY* gene family in *Populus trichocarpa*. *Plant Cell Rep*. 2012;31:1199–217.
- Mao K, Hayashi S, Kojima-Suzuki H, Morikami A, Nakamura K. Role of conserved residues of the WRKY domain in the DNA-binding of tobacco WRKY family proteins. *Biosci Biotechnol Biochem*. 2001;65:2428–36.
- Ciolkowski I, Wanke D, Birkenbihl RP, Somssich IE. Studies on DNA-binding selectivity of WRKY transcription factors lend structural clues into WRKY-domain function. *Plant Mol Biol*. 2008;68:81–92.
- Rushton PJ, Somssich IE, Ringler P, Shen QJ. WRKY transcription factors. *Trends Plant Sci*. 2010;15:247–58.
- Sun C, Palmqvist S, Olsson H, Borén M, Ahlandsberg S, Jansson C. A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the *iso1* promoter. *Plant Cell*. 2003;15:2076–92.
- Chen S, Li J, Wang S, Fritz E, Hüttermann A, Altman A. Effects of NaCl on shoot growth, transpiration, ion compartmentation, and transport in regenerated plants of *Populus euphratica* and *Populus tomentosa*. *Can J For Res*. 2003;33:967–75.
- Jiang Y, Deyholos MK. Comprehensive transcriptional profiling of NaCl-stressed *Arabidopsis* roots reveals novel classes of responsive genes. *BMC Plant Biol*. 2006;6:25.
- Qiu Y, Yu D. Over-expression of the stress-induced *OsWRKY45* enhances disease resistance and drought tolerance in *Arabidopsis*. *Environ Exp Bot*. 2009;65:35–47.
- Shi W, Liu D, Hao L, Wu C-A, Guo X, Li H. *GbWRKY39*, a member of the WRKY transcription factor family in cotton, has a positive role in disease resistance and salt stress tolerance. *Plant Cell Tissue Organ Cult*. 2014;118:17–32.
- Cai R, Zhao Y, Wang Y, et al. Overexpression of a maize *WRKY58* gene enhances drought and salt tolerance in transgenic rice. *Plant Cell Tissue Organ Cult*. 2014;119:565–77.
- Jiang Y, Deyholos MK. Functional characterization of *Arabidopsis* NaCl-inducible *WRKY25* and *WRKY33* transcription factors in abiotic stresses. *Plant Mol Biol*. 2009;69:91–105.
- Shen Z, Yao J, Sun J, et al. *Populus euphratica* HSF binds the promoter of *WRKY1* to enhance salt tolerance. *Plant Sci*. 2015;235:89–100.
- Browicz K. Chorology of *Populus euphratica* Olivier. *Arboretum Kornickie*. 1977;22:5–27.
- Chen S, Li J, Fritz E, Wang S, Hüttermann A. Sodium and chloride distribution in roots and transport in three poplar genotypes under increasing NaCl stress. *For Ecol Manage*. 2002;168:217–30.
- Ottow EA, Brinker M, Teichmann T, et al. *Populus euphratica* displays apoplastic sodium accumulation, osmotic adjustment by decreases in calcium and soluble carbohydrates, and develops leaf succulence under salt stress. *Plant Physiol*. 2005;139:1762–72.
- Ma T, Wang J, Zhou G, et al. Genomic insights into salt adaptation in a desert poplar. *Nat Commun*. 2013;4:2797.
- Tuskan GA, Difazio S, Jansson S, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 2006;313:1596–604.
- Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14:755–63.
- Finn RD, Mistry J, Tate J, et al. The Pfam protein families database. *Nucleic Acids Res*. 2010;38:D211–22.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;1:9.
- Wang Y, Tang H, Debarry JD, et al. MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40:e49–9.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22:4673–80.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9.
- Bailey TL, Boden M, Buske FA, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*. 2009;37:W202–8.
- Overbeek R, Fonstein M, D'souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci U S A*. 1999;96:2896–901.
- Zhang J, Xie P, Lascoux M, Meagher TR, Liu J. Rapidly evolving genes and stress adaptation of two desert poplars, *Populus euphratica* and *P. pruinosa*. *PLoS One*. 2013;8:e66370.
- Löytynoja A, Goldman N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A*. 2005;102:10557–62.
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5:621–8.



45. Zhang Y, Mao L, Wang H, et al. Genome-wide identification and analysis of grape aldehyde dehydrogenase (ALDH) gene superfamily. *PLoS One*. 2012;7:e32153.
46. Park CY, Lee JH, Yoo JH, et al. WRKY group IId transcription factors interact with calmodulin. *FEBS Lett*. 2005;579:1545–50.
47. Zhang Y, Wang L. The WRKY transcription factor superfamily: its origin in eukaryotes and expansion in plants. *BMC Evol Biol*. 2005;5:1.
48. Wang M, Vannozzi A, Wang G, et al. Genome and transcriptome analysis of the grapevine (*Vitis vinifera* L.) WRKY gene family. *Hortic Res*. 2014;16:1.
49. Jiang Y, Duan Y, Yin J, et al. Genome-wide identification and characterization of the *Populus* WRKY transcription factor family and analysis of their expression in response to biotic and abiotic stresses. *J Exp Bot*. 2014;65:6629–44.
50. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol*. 2000;17:32–43.
51. Journot-Catalino N, Somssich IE, Roby D, Kroj T. The transcription factors *WRKY11* and *WRKY17* act as negative regulators of basal resistance in *Arabidopsis thaliana*. *Plant Cell*. 2006;18:3289–302.
52. Gao Q-M, Venugopal S, Navarre D, Kachroo A. Low oleic acid-derived repression of jasmonic acid-inducible defense responses requires the *WRKY50* and *WRKY51* proteins. *Plant Physiol*. 2011;155:464–76.
53. Yuan S, Hu Q. The function of jasmonic acid in the salt tolerance of *Populus euphratica*. *J Beijing Forestry Univ*. 2002;24:66–9.
54. Besseau S, Li J, Palva ET. *WRKY54* and *WRKY70* co-operate as negative regulators of leaf senescence in *Arabidopsis thaliana*. *J Exp Bot*. 2012;63:2667–79.