J. For. Res.
https://doi.org/10.1007/s11676-018-0734-2

ORIGINAL PAPER

Genome-wide identification and characterization of *WUSCHEL*-related homeobox (*WOX*) genes in *Salix suchowensis*

Xuelin Wang[1] · Changwei Bi[2] · Chunyan Wang[1] · Qiaolin Ye[1] · Tongming Yin[3] · Ning Ye[1]

Received: 23 April 2017 / Accepted: 29 August 2017

Abstract Members of the *WUSCHEL*-related homeobox (WOX) transcription factor family are essential for determining cell fate and regulating diverse developmental processes in plants. Many *WOX* genes have been systematically investigated in woody plants such as *Populus trichocarpa*, but not in *Salix suchowensis*. Whole-genome sequence data for *S. suchowensis* is now available for comprehensive study of *WOX* genes in *S. suchowensis*. We thus surveyed the genome of *S. suchowensis* and demonstrated active expression of 15 *WOX* genes. In a phylogenetic analysis of *WOX* genes, the 15 *SsWOX* genes clustered among the modern/WUS, intermediate and ancient clades similar to the *WOX* genes of *Arabidopsis thaliana*. Based on the conserved intron/exon structure, *SsWOX* genes in the same subgroup had similar conserved exon–intron structures and motif domains. Furthermore, among several *SsWOX* subgroups, WUS (Wuschel)-box and EAR (the ERF-associated amphiphilic repression)-like motifs were conserved. Expression profiles of *WOX* genes in roots, stems and leaves indicate that *SsWOX* genes have various conserved roles in the tissues. Comparative analysis of the expression patterns in *Salix suchowensis* with that of *Arabidopsis* suggests that different shoot regeneration abilities are controlled by different *WOX* genes in plants. The analysis provide an overview of differentially expressed *SsWOX* genes during shoot regeneration, but also contribute to understanding the evolution of *WOX* genes in Salicaceae and the interrelations of *WOX* genes and other transcription factors, providing targets for further study.

Keywords WOX family · Salicaceae · Expression · Evolution · Duplication

Project funding: This study was supported by the National Key Research and Development Plan of China (2016YFD0600101), the Fundamental Research Funds for the Central Non-Profit Research Institution of CAF (CAFYBB2014QB015), the National Natural Science Foundation of China (31570662, 31500533, and 61401214), Jiangsu Provincial Department of Housing and Urban–Rural Development (2016ZD44), 2017 Graduate Research and Innovation Program Projects in Jiangsu Province (KYCY17_0827), and the PAPD (Priority Academic Program Development) program at Nanjing Forestry University.

The online version is available at http://www.springerlink.com

Corresponding editor: Yu Lei.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11676-018-0734-2) contains supplementary material, which is available to authorized users.

✉ Ning Ye
yening@njfu.edu.cn

[1] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, People's Republic of China

[2] School of Biological Science and Medical Engineering, Southeast University, Nanjing 210037, People's Republic of China

[3] College of Forest Resources and Environment, Nanjing Forestry University, Nanjing 210037, People's Republic of China

Introduction

WUSCHEL-related homeobox (WOX) transcription factors, a subgroup of the HB transcription factors, have a conserved helix-loop-helix-turn-helix structure and can control cell fate and differentiation (Cao et al. 2017). WOXs in plants are separated on the basis of the phylogenetic relatedness of the homeodomain. WOXs play significant roles in developmental and growth processes including embryonic patterning, stem cell maintenance and organ formation (Dolzblasz et al. 2016). The function and evolution of *WOX* genes in *Arabidopsis thaliana* have been well studied and characterized (Jin et al. 2017), and recently some WOX members of *Arabidopsis* were shown to be essential for organ formation, stem cell maintenance and embryonic patterning (Laux et al. 1996; Haecker et al. 2004; Graaff et al. 2009). *Arabidopsis* WOX proteins also regulate lateral organ development, embryo apical-basal polarity patterning and stem cell maintenance in the shoot apical meristem (SAM), root apical meristem (RAM) and cambium (CAM; Lin et al. 2013). Because of the important role of *WOX* genes in cell control and the recent availability of the full genomic sequence for the shrub willow *Salix suchowensis* (Dai et al. 2014), we characterized the *WOX* genes in *S. suchowensis*.

WOX genes have been divided into three clades, the modern, the intermediate and the ancient clade (Liu et al. 2014). *WOX* genes in green algae and nonvascular mosses including the model moss *Physcomitrella patens* are found only in the ancient clade, whereas *WOX* genes of rice, sorghum and maize are found in all three clades (Zhang et al. 2010). For *Arabidopsis*, *WUS* and *AtWOX1-7* belong to the modern *WUS* clade; *AtWOX8*, *-9*, *-11* and *-12* belong to the intermediate clade; and *AtWOX10*, *-13* and *-14* group with the ancient clade (Graaff et al. 2009; Zhang et al. 2010). Among the modern *WOX* genes, *Arabidopsis WUSCHEL* (*AtWUS*) is required, with CLV3 through a feedback loop, for the vegetative-to-embryonic transition and maintaining the shoot apical meristem together. The expression of *AtWUS* is also especially abundant in the organizing center in the SAM. When overexpressed, *AtWOX1* disturbs usual meristem growth (Zhang et al. 2011). *AtWOX2*, which participates in determining cell fate of apical and basal cell lineages during embryonic development, is expressed in zygotes (Breuninger et al. 2008). By employing organ founder cells and forming the lateral domain, *AtWOX3 (PRS1)* contributes to the lateral development of floral and vegetative organs (Shimizu et al. 2009). Its function in controlling lateral organ development can also be completely replaced by *AtWUS* and partly replaced by *AtWOX4* (Ji et al. 2010a). *AtWOX4* is highly expressed to control the maintenance of vascular stem cells in the CAM (Hirakawa et al. 2010; Suer et al. 2011). *AtWOX5* is extremely important in the maintenance of stem cells through a negative feedback signal via CLE40 (Oshchepkova et al. 2017), and *AtWOX6* controls ovule development (Yang et al. 2017). With respect to the maintenance of stem cells in the SAM and RAM, *AtWUS* and *AtWOX5* are functionally interchangeable (Sarkar et al. 2007). Of the intermediate and ancient *WOX* genes, the modern *AtWOX2* in the zygote, cooperates with *AtWOX8* in pre-embryo development (Breuninger et al. 2008). In the *Arabidopsis* SAM, *AtWOX9* takes part in differentiation by preventing premature cell division and maintaining appropriate timing of cell division (Yang et al. 2017). The function and expression of the most conserved ancient WOX proteins, *AtWOX10*, are still unknown. *AtWOX11* and *-12* participate in root organogenesis, *AtWOX13* facilitates replum formation during fruit development and is highly expressed in developing flowers and young siliques (Romera-Branchat et al. 2013). Like *AtWOX4* in the modern WOX clade, *AtWOX14* has the same function in controlling vascular meristem development. As a consequence, the different functions of the different *WOX* geneswill be hot topics in the future (Hirakawa et al. 2010; Ji et al. 2010a, b; Etchells et al. 2013; Luan et al. 2013).

Salix suchowensis has wide uses worldwide, and its growth process can have a direct impact on its quality. However, the roles of *WOX* genes in *S. suchowensis*, especially how *WOX* genes influence the regulation of shoot regeneration are unknown. The availability of the sequenced genome provides an opportunity for genome-wide identification and analysis of *WOX* genes. We, therefore, conducted a comprehensive investigation on *WOX* genes in *S. suchowensis*; the study including identification of WOX-encoding sequences, phylogenetic analysis, chromosome location, exon/intron structure, gene duplication events, conserved motif and expression pattern. The results will provide an overview for the transcription regulation of *S. suchowensis* shoot regeneration and serve as a guideline for the future.

Materials and methods

Database search for *WOX* genes

RNA extraction used five tissues including tend root, young leaves, bark, non-lignified shoot and vegetative buds of *S. suchowensis* and RNA concentration was measured using a Nanodrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA; Dai et al. 2014). Then, DNA residues were digested at 37 °C for 30 min by DNase (TaKaRa, Shiga, Japan), RNA was purified using Oligotex mRNA Mini Kit (Qiagen, Duesseldorf, Germany).

Furthermore, cDNA was synthesized using a cDNA Synthesis System Kit (Roche, Basel, Switzerland). Finally, the genome of *S. suchowensis* was sequenced using Roche/454 and Illumina/HiSeq-2000 sequencing technologies. Genome annotation and predicted protein sequences for *S. suchowensis* are accessible using version5_2.gff3 and Willow.gene.pep, respectively, at website http://bio.njfu.edu.cn/ss_wox/.

Putative *WOX* genes of *S. suchowensis* were identified using a method similar to that used for other species (Bi et al. 2016). The Pfam database (http://pfam.xfam.org/) supplies the Hidden Markov Model (HMM) profile of SsWOX transcription factors and was downloaded using the key ID 'PF00046' (Finn et al. 2016). To find *SsWOX* gene sequences in Willow.gene.pep, we used the HMM profile as a query in the BLASTP program with E-value cutoff set to 1e-5 (Zhang et al. 2015). To validate the accuracy of the search, we downloaded WOX protein sequences for *Arabidopsis* from TAIR (http://www.arabidopsis.org) and for *Populus trichocarpa* from PlantTEDB (http://planttfdb.cbi.pku.edu.cn/family.php?fam=WOX).

The HMMER program (hmmbuild) built a HMM model using the WOX protein sequence alignment of *Arabidopsis* and *P. trichocarpa* in the Stockholm format, then the model was used to search *S. suchowensis* WOX protein sequences by HMMER program (hmmsearch; Ge et al. 2016). Finally, conserved SsWOX HD domains were confirmed using the programs SMART (http://smart.embl-heidelberg.de/) and ClustalW (version 2.1). Boxshade (http://www.ch.embnet.org/software/BOX_form.html) was used to color and indicate these domain sequences.

SsWOX genes were mapped onto chromosomes (http://bio.njfu.edu.cn/ss_wrky/version5_2.fa) using an in-house Perl script (http://bio.njfu.edu.cn/willow_chromosome/BuildGff3_Chr.pl). The isoelectric point (PI) and molecular weight (MW) of the predicted protein encoded by each putative *WOX* gene was calculated using ExPasy site (https://web.expasy.org/compute_pi/). Mapping of each *WOX* gene to the chromosomes was depicted by MapInspect software (http://mapinspect.software.informer.com/).

Phylogenetic analysis and classification of the *WOX* genes in *Salix suchowensis*

Fifteen SsWOX and 15 AtWOX proteins sequences were aligned using ClustalW and default parameters to achieve a better classification of the *SsWOX* genes (Larkin et al. 2007). Only HD regions of selected plants were aligned and analyzed for the phylogenetic analysis (Larkin et al. 2007). The HD region and complete predicted amino acid sequences were used to construct a phylogenetic tree in MEGA7 using the neighbor-joining (NJ) method (Kumar et al. 2016). Bootstrap values calculated from 1000 iterations in the pairwise gap deletion mode is helpful for determining the topology of the NJ tree and inferring relationships among species (Bi et al. 2016). A more detailed sequence alignment of all 15 SsWOXs domains, 10 AtWOXs domains, 18 PtrWOXs domains, 11 *O. galberrima* WOXs domains, 7 *C. lanatus* WOXs domains and 10 *V. vinifera* WOXs domains was performed to compare WOXs from different species in different families.

Exon/intron structure, gene duplication events and conserved motif distribution of *WOX* genes in *Salix suchowensis*

The protein annotation file (http://bio.njfu.edu.cn/ss_wrky/version5_2.gff3) and the online gene structure display server (GSDS: http://gsds.cbi.pku.edu.cn/) were used to obtain the exon/intron structures of *WOX* genes of *S. suchowensis* (Bailey et al. 2006, 2015).

Gene duplication was detected in our study. Gene duplication discovery was executed using BLASTP with E-value cutoff set to 1e-20 and the following requirements (Gu et al. 2002; Bi et al. 2016): (1) The percentage of aligned sequence coverage in the long gene sequence should exceed 80%; (2) the percentage of aligned area similarity should exceed 70%.

Multiple expectation maximization for motif elicitation (MEME), an online tool (http://meme-suite.org/tools/meme), was used to find conserved motifs and identify structural features of predicted protein sequences for the *SsWOX* genes. HD regions motifs were interpreted using the http://meme-suite.org/tools/meme website. We used the following settings to run the MEME: the optimum motif widths ranged from 6 to 21 residues, maximum number of motifs was 20 and number of repetitions was "any" (Bailey et al. 2006).

Expression of *WOX* genes in *Salix suchowensis*

To better understand the contributions of *SsWOX* genes to development and physiological functions, *S. suchowensis* RNA-HiSeq sequenced reads from three tissues (roots, stems and leaves) were used to quantify expression of *WOX* genes. Using the Burrows–Wheeler alignment (BWA; Li and Durbin 2009) with mismatch $\leq$ 2 bp and other default settings, we mapped back onto the respective SsWOX protein sequences and counted the numbers of mapped reads for each *SsWOX* gene. Mapped reads were normalized for sequencing depth and gene length by calculating the number of reads per kilo base per million reads (RPKM). R package (R Core Team, 2004–2016, GUI 1.69) and the log2 RPKM values were used to draw a heat map to profile tissue-specific expression of *SsWOX* genes in all tissues sampled (Bi et al. 2016).

Results

Identification and characterization of 15 WOX proteins in *Salix suchowensis*

To initiate analysis of WOX proteins in *S. suchowensis*, we used HMMER in protein sequences to explore the HMM profile of the WOX DNA-binding domain, then used BLASTP to confirmed the accuracy of the results. Fifteen known WOX protein sequences of *Arabidopsis* were used as queries in a BLASTP search of the *S. suchowensis* genomic database. We obtained 18 putative WOX proteins and further validated them using the online SMART program. As a consequence, three proteins that lacked a complete WOX domain were removed, and 15 probable WOX proteins were chosen as WOX superfamily members, close to the number of WOX proteins found in *P. trichocarpa* (18), *A. thaliana* (15), *Oryza sativa* (13), and *Vitis vinifera* (12).

Similar to the known clade groupings of WOX proteins from other plants, the 15 SsWOX proteins fall into three major clades on the basis of the HD sequence: the modern/WUS clade, the intermediate clade, and the ancient clade, similar to the clustering of their *Arabidopsis* counterparts. According to the *Arabidopsis* counterparts in *S. suchowensis*, the modern/WUS clade has eight SsWOXs, which group into six subclasses: SsWUSA, SsWUSB, SsWOX1A, SsWOX1B, SsWOX2, SsWOX4, SsWOX5 and SsWOX6. The intermediate clade was divided into three subclasses and consists of four SsWOXs: SsWOX9, SsWOX11 and SsWOX12. The ancient clade possesses one subclass, with three SsWOXs: SsWOX13 (SsWOX13A, SsWOX13B and SsWOX13C). All the SsWOXs were checked against the exon/intron organization of WOX-coding sequences of *Arabidopsis*.

Figure 1 suggests that all SsWOX proteins have a highly conserved helix-loop-helix-turn-helix domain, and most WOX proteins contain the complete or incomplete WUS box. The HD area includes several conserved amino acids such as Q, L and Y in helix 1, P, I, L in helix 2, N, V, W, F, Q, N, R and R in helix 3 and G in turn (Fig. 1). Apart from the conserved amino acid residues, we found additional conserved residues, including I in helix 2, F, Y in helix 3 among these WOX members.

Two copies of SsWOXs were classified as WUS (SsWUSA and B), WOX1 (SsWOX1A and B) and WOX9 (SsWOX9A and B), and three copies of SsWOXs could be classified as WOX13 (SsWOX13A, B and C). Nevertheless, no SsWOXs were closely related to AtWOX3, AtWOX7, AtWOX8, AtWOX10, or AtWOX14 in *Arabidopsis*. Similarly, close orthologs of AtWOX7, AtWOX10, and AtWOX14 are not present in *P. trichocarpa* (Liu et al. 2014), *Vitis vinifera*, *Picea abies* and other woody plant species.

Chromosomal location of *WOX* genes in *Salix suchowensis*

The location of *WOX* genes on the chromosome provides a better understanding of how *SsWOX* genes expand in the sequenced genome of *S. suchowensis.* Fourteen of the fifteen putative *SsWOX* genes could be mapped onto one or two ends of 11 chromosomes, while no *WOX* genes of *S. suchowensis* were located on the other 8 chromosomes. Most of the 15 *SsWOX* genes were distributed evenly among the 11 chromosomes of *S. suchowensis*, with the exception of chromosome V and X. One *SsWOX* gene (sequence ID: willow_GLEAN_10001899), labeled as *SsWOX1B*, could not be definitively mapped onto any chromosome. Figure 2 provides a detailed diagram of the *WOX* gene location on the chromosomes of *S. suchowensis*. Chromosome (Chr) 5 has the most *SsWOX* genes (3), followed by Chr10 (2). One *SsWOX* gene was identified on Chr2, Chr7, Chr9, Chr11, Chr12, Chr13, Chr14, Chr17, Chr19 and Chr20; Chr1, Chr3, Chr4, Chr6, Chr8, Chr15, Chr16 and Chr18 had none.

Characteristics and functional domains of *SsWOX* genes

Detailed characteristics including *WOX* genes sequence ID, specific chromosomal distribution, specific group attribution, amino acid lengths, PI/MW and introns of *SsWOX* genes are given in Table 1. As is shown in Table 1, compared to WOX protein sequences of *P. trichocarpa*, which comprise from 171 to 390 amino acids (Liu et al. 2014), the *SsWOX*-encoded protein sequences ranged from 213 residues (SsWOX1A) to 450 residues (SsWOX5), for an average length of 279 residues. In addition, the PI varied from 5.13 (SsWOX13C) to 9.20 (SsWOX9B), and the MW varied from 24.26 (SsWUSB) to 50.94 kDa (SsWOX5).

The *Arabidopsis* WUS (AtWUS) protein contains three functional domains: the acidic region, WUS box and EAR-like motif, which contribute significantly to its transcription factor function (Ikeda et al. 2009). Downstream of the HD domain is a conserved WUS-box domain (TLXLFP), which was found in the modern clade members: SsWUS, SsWOX2, SsWOX4, SsWOX5 and SsWOX6, but not in the other clades. Furthermore, there is an EAR-like domain in the C-terminal ends of SsWUSA (Fig. 3).

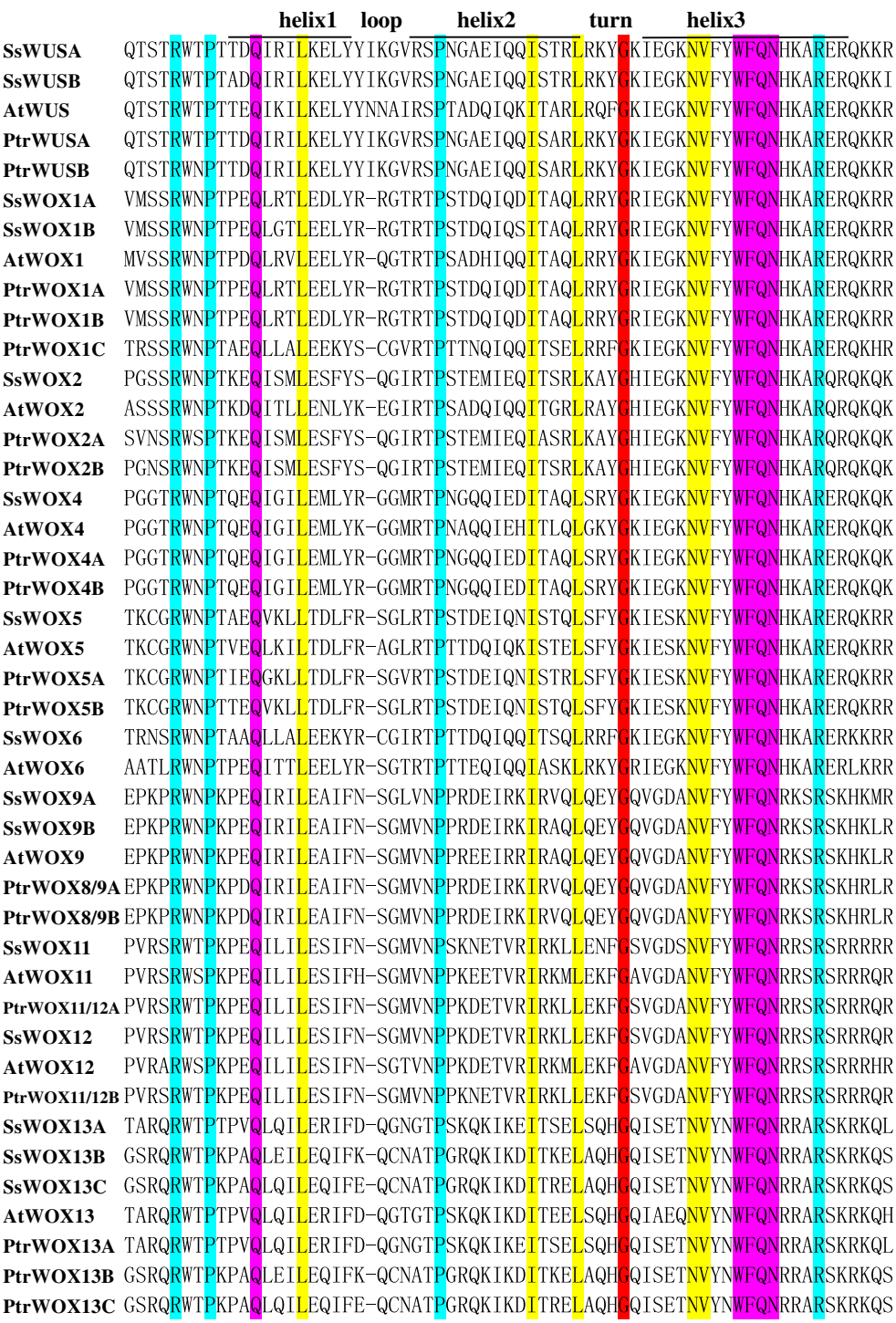

Fig. 1 Alignment of the WOX homeodomain sequences. The shaded areas indicate several highly conserved residues R, P, Q, L, P, I, L, G, N, V, W, F, Q, N, R by alignment of the homeodomains in the five angiosperms

Phylogenetic analysis of WOX proteins in *Arabidopsis,* Salicaceae crops and four other eudicots

The phylogenetic analysis of WOXs was based on 70 amino acid sequences using the NJ method. To further study the evolutionary relationships, we built a phylogenetic tree based on WOX proteins sequences of *S. suchowensis, A. thaliana* and other four eudicot plants including *P. trichocarpa*, *Oryza galberrima*, *Citrullus lanatus* and *Vitis vinifera* (Fig. 4). We chose AtWOXs as the model system (Graaff et al. 2009; Zhang et al. 2010). Furthermore, a phylogenetic tree could also be constructed by protein-coding genes of plant organellar genomes (Wang et al. 2018; Ye et al. 2017).

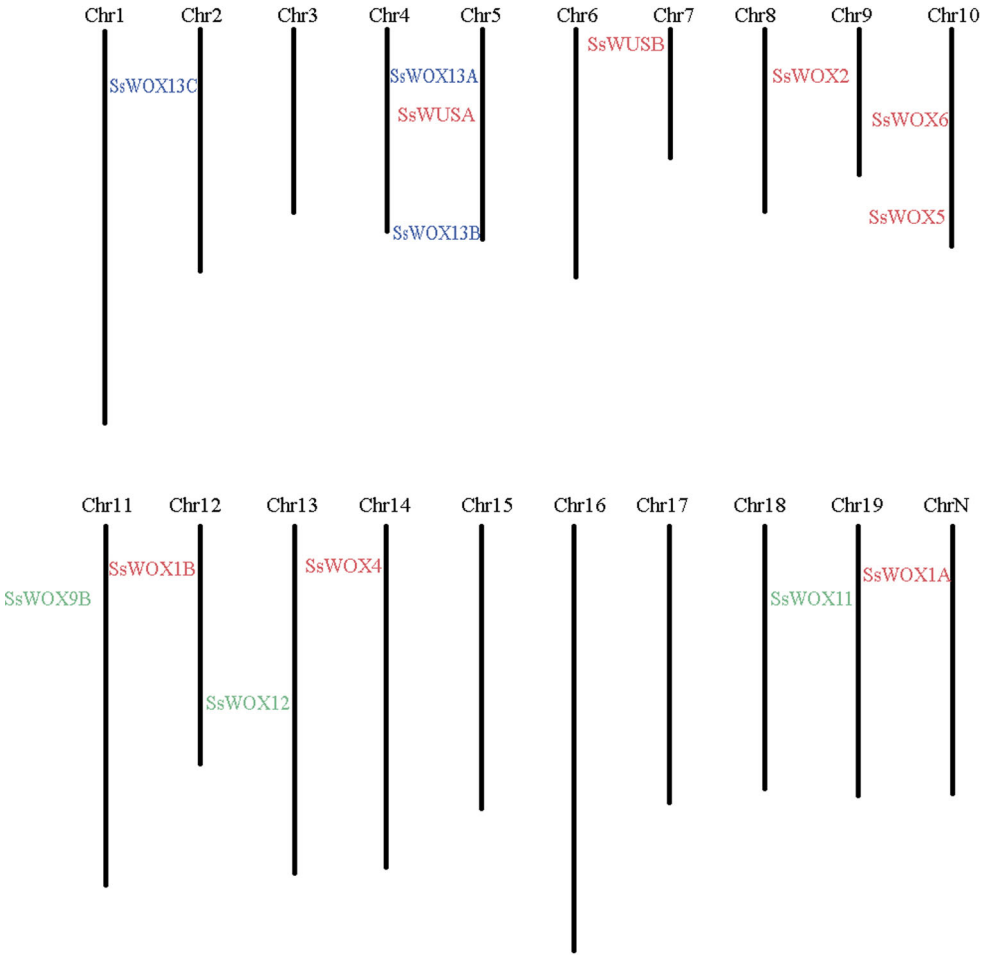

Fig. 2 Chromosomal location of *SsWOX* genes. Chromosome number is above each chromosome; left side of each chromosome is related to the approximate physical location of each *WOX* gene. Only one unmapped *SsWOX1A* gene is shown on ChrN. Red: modern group, green: intermediate group, blue: ancient group

Table 1 Characteristics of WOX proteins identified in *S. suchowensis*

SsWOXs	Sequence ID	Chr	Group	Length (aa)	Deduced polypeptide		Introns
					PI	MW (kDa)	
SsWUSA	willow_GLEAN_10004905	5	I	265	5.98	30.04	2
SsWUSB	willow_GLEAN_10012646	7	I	216	6.52	24.26	2
SsWOX1A	willow_GLEAN_10001899	N/A	I	213	8.99	24.72	2
SsWOX1B	willow_GLEAN_10019781	12	I	378	6.18	42.43	3
SsWOX2	willow_GLEAN_10022876	9	I	237	7.15	27.13	1
SsWOX4	willow_GLEAN_10004293	14	I	216	8.73	24.85	2
SsWOX5	willow_GLEAN_10019395	10	I	450	6.26	50.94	11
SsWOX6	willow_GLEAN_10021393	10	I	295	8.37	33.54	3
SsWOX9A	willow_GLEAN_10007329	17	II	393	7.28	43.11	4
SsWOX9B	willow_GLEAN_10007682	11	II	318	9.20	35.63	1
SsWOX11	willow_GLEAN_10004422	19	II	236	5.67	25.92	1
SsWOX12	willow_GLEAN_10011530	13	II	269	6.16	29.77	2
SsWOX13A	willow_GLEAN_10011009	5	III	273	5.94	30.95	3
SsWOX13B	willow_GLEAN_10017952	5	III	217	5.58	24.76	2
SsWOX13C	willow_GLEAN_10022640	2	III	215	5.13	24.56	2

Chr chromosome numbers, *N/A* not available, *M* (I) modern clade, *I* (II) intermediate clade, *A*(III) ancient clade

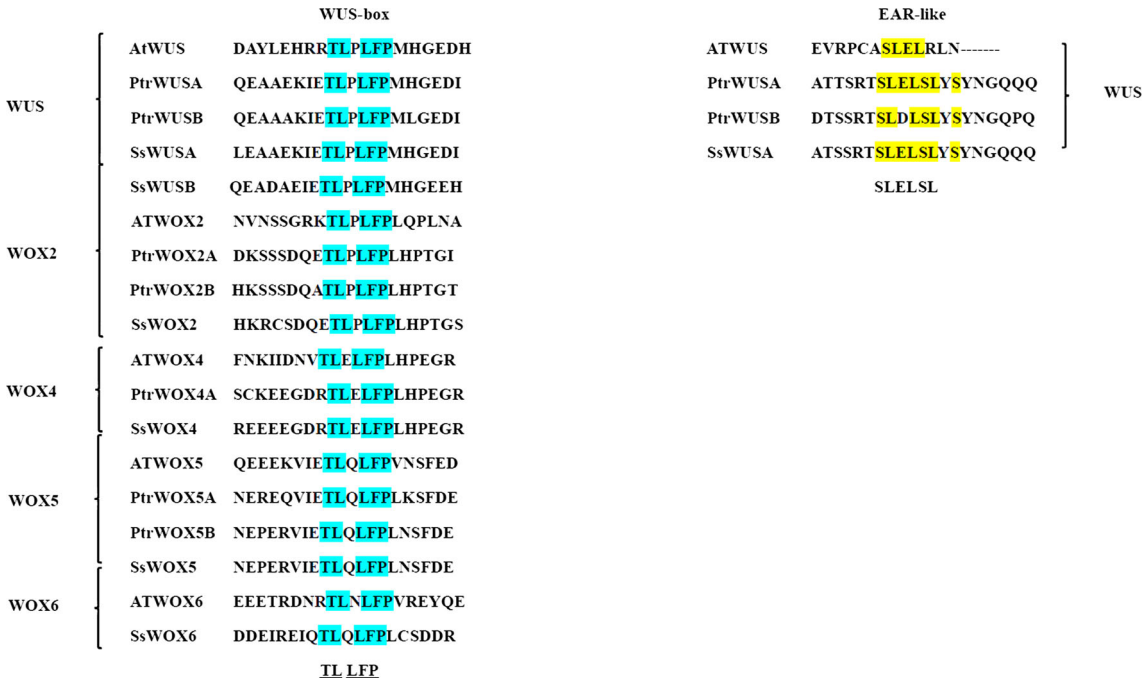

Fig. 3 Multiple sequence alignment of WUS-box and EAR-like domains of WOXs in *A. thaliana*, *P. trichocarpa*, and *S. suchowensis*. Left: WUS boxes from WUS homologs and WOX proteins; the 18 WOXs had WUS boxes and belonged to the modern clade. Conserved residues are highlighted and shown below the alignment. Right: EAR-like domains from WUS and WOX5 proteins. WOX proteins of subgroups WUS and WOX5 contain this functional domain. Conserved residues are highlighted and shown below the alignment

Comparatively well-supported phylogenetic trees based on the entire WOX protein sequences and WOX domain protein sequences were almost identical, even though some bootstrap values of branches were relatively low. Therefore, we used the WOX HD sequences to construct the phylogenetic tree. The 70 WOX members were allocated into three clades. The modern clade (also called the first clade) contained 39 WOXs, including 8 from *S. suchowensis*, 6 from *A. thaliana*, 11 from *P. trichocarpa*, 6 from *C. lanatus*, 4 from *O. galberrima*, and 4 from *V. vinifera*, which are homologous to *Arabidopsis* WUS and WOX1, WOX2, WOX4, WOX5 and WOX6 (Lian and Ding 2014). The intermediate clade (also called the second clade) consisted of 17 WOXs, which are homologous to *Arabidopsis* WOX8, WOX9, WOX11, and WOX12; 6 WOXs were from *O. galberrima*, 4 from *S. suchowensis*, 4 from *P. trichocarpa*, 3 from *A. thaliana*, 2 from *V. vinifera* and just 1 from *C. lanatus*. These members in the intermediate clade were further divided into two subgroups: WOX8/9 and WOX11/12. More specifically, WOX8/9 contained 10 members and WOX11/12 had 10 members. The third clade (also called the ancient Clade) included 11 WOXs (Fig. 4; Lian and Ding 2014).

To achieve a better study in model plant and woody plant, a phylogenetic tree was constructed based on the WOX domains of *A. thaliana*, *S. suchowensis* and *P. trichocarpa* (Fig. S1). The tree depicts that the majority of the WOX domains from *S. suchowensis* and *P. trichocarpa* are clustered into sister. Furthermore, a total of five SsWOX domains show extremely the same domains (similarity: 100%) to *P. trichocarpa*: SsWOX1A and PtrWOX1B, SsWOX4 and PtrWOX4A, SsWOX4 and PtrWOX4B, SsWOX12 and PtrWOX11/12A, SsWOX13C and PtrWOX13C. Subsequent functional analysis of these proteins in *S. suchowensis* and *P. trichocarpa* would provide a useful reference for other species in Salicaceae.

In further studies of WOX protein sequences of *P. trichocarpa* and *S. suchowensis*, we demonstrated that the sequences are greatly conserved. The HD region possesses some conserved amino acids of the two Salicaceae species, e.g., Q and L in helix 1; P, I and L in helix 2; I, N, V, W, F, Q, N, and R in helix 3; and G in turn (T) (Fig. 1). Furthermore, we found that in the modern clade a great number of WOX protein sequences contained a complete WUS box (amino acids, TL-LFP) although AtWOX7, SsWOX1A, VvWOX3, and ClaWOX6 did not. In contrast, WOX protein sequences in the intermediate and ancient clade had just one conserved amino acid, F, as for the WUS box motif.

Exon/introns of *SsWOX* genes

In plant evolution, exon–intron structures are significant for most genes families. The phylogenetic relationships of

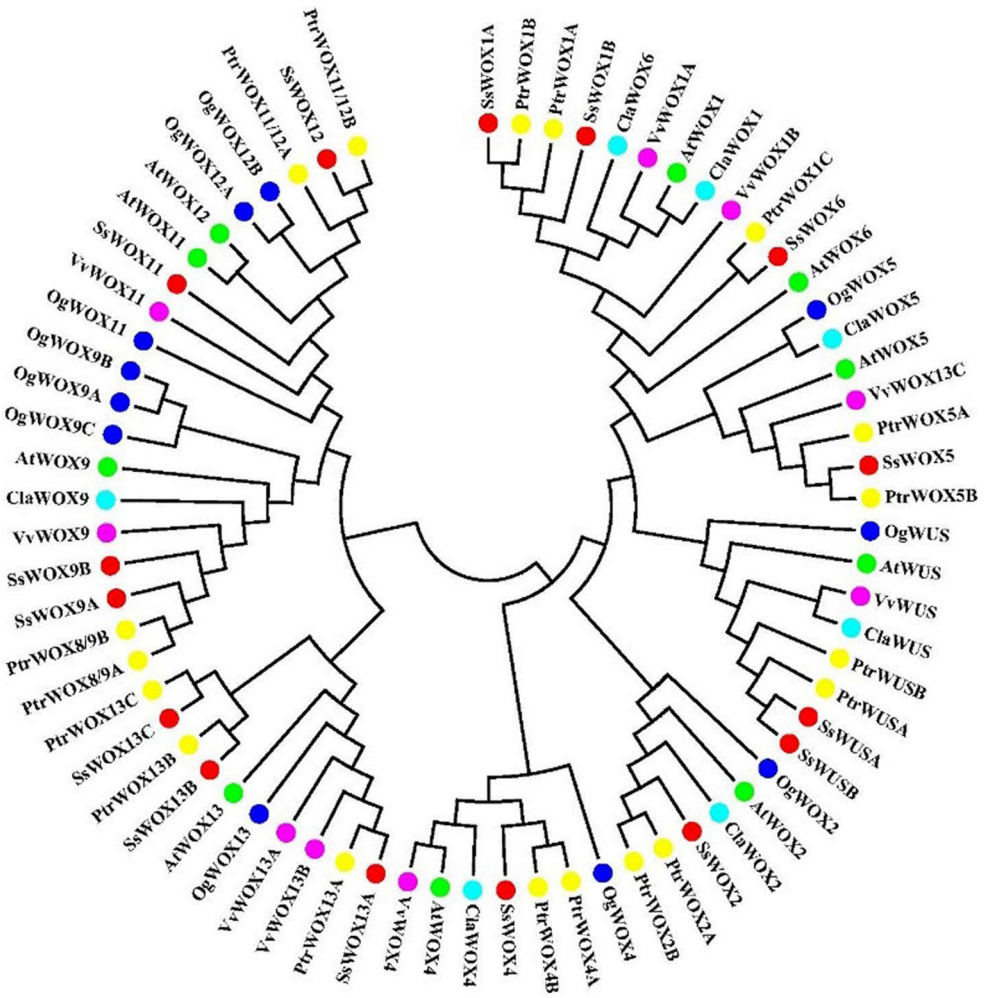

Fig. 4 Phylogenetic tree based on WOX domains from proteins in *S. suchowensis, A. thaliana*, *P. trichocarpa*, *O. galberrima*, *C. lanatus* and *V. vinifera* and constructed using the neighbor-joining method in MEGA7. Circles with different colors indicate different species

multiple gene families can be represented by positions, phases and length of its introns. Therefore, using coding sequences and genomic sequences through GSDS, we investigated exon/intron positions and phases of the 15 *WOX* genes in *S. suchowensis*. Detailed information on length and phases of introns, corresponding exon sequences and phylogenetic relationships of each *SsWOX* gene are in Fig. 5. With the exception of *SsWOX1A*, *SsWOX5*, *SsWOX9*, *SsWOX12* and *SsWOX13A*, other *SsWOX* genes have a similar intron/exon organization and size to the WOX genes of *Arabidopsis* counterparts in order and size (Zhang et al. 2010). Most *SsWOX* genes have 1–3 introns: one intron for *SsWOX2*, *-9B* and *-11*; two introns for *SsWUSA*, *SsWUSB*, *SsWOX1A*, *SsWOX4*, *-12*, *-13B* and -

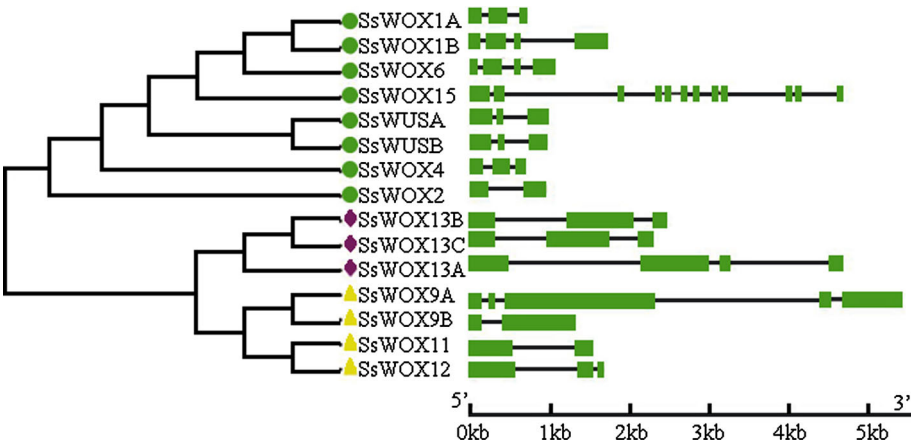

Fig. 5 Phylogenetic tree and genomic organization of *SsWOX* genes. Left, phylogenetic tree built using the full-length *SsWOX* and the neighborhood joining method in MEGA7. Right, exon/intron structure of *SsWOX* genes displayed using GSDS. Green: exons; gray, introns. Introns 0, 1 and 2 are indicated by 0, 1 and 2, respectively

13C; and three introns for *SsWOX1B*, *-6* and *-13A*. *WOX4* gene from *Arabidopsis*, *S. suchowensis* and *P. trichocarpa* each contain similar intron phases and exon patterns. Furthermore, in spite of the variations in the intron phases, *SsWOX1B* and *AtWOX1* almost have the same gene structures. Interestingly, the length of the second intron of *SsWOX5* is more than 2 kb, and the size of third intron *SsWOX9A* is more than 2 kb (Fig. 5), which has not been found for *WOX* genes of *Arabidopsis* and *P. trichocarpa*.

Identification of gene duplication events and conserved motifs in *Salix suchowensis*

In the duplicated segments of the *S. suchowensis* genome, we discovered four pairs of *SsWOX* genes, *SsWUSA/SsWUSB*, *SsWOX1A/SsWOX1B, SsWOX9A/SsWOX9B* and *SsWOX13B/SsWOX13C*. The percentage of gene duplication events in three clades in ascending order are the modern clade: 50% (4 of 8 have gene duplication events), the intermediate clade: 50% (2 of 4 have gene duplication events) and the ancient clade: 66.7% (2 of 3 have gene duplication events). None of the four homologous gene pairs (*SsWOX13A/SsWUSA*, *SsWOX13A/SsWOX13B*, *SsWUSA/SsWOX13b*, *SsWOX5/SsWOX6*) have undergone TDs. On the contrary, all 4 pairs of *SsWOX* genes, which account for 53.3% of all *SsWOX* genes, have participated in SDs. Consequently, chromosomal duplication events may have led to the expansion of the *SsWOX* gene during the evolution of Salicaceae.

In WOX proteins, conserved core domains with about 60–66 residues regulate more functions, whereas other sequences control a few (Yang et al. 2017). Here, we used the online program MEME to predict the conserved motifs and to obtain a better understanding of the structural features of proteins encoded by *WOX* genes of *S. suchowensis*. Among 20 putative motifs, motif 1, 2 and 5 were present at high frequency in *SsWOX* genes and have conserved domain characteristics of WOX proteins. Motif 4, 6, and 16 were found in the ancient clade; motif 3, 7, 9, 10, and 14 were only found in the intermediate clade; motif 5, 8, 11, 12, 15, and 17–20 were found in the modern clade (Fig. S2). Thus, motifs in the WOX gene family will be a crucial foundation for future functional and structural studies.

There are three functional domains of *Arabidopsis* WUS (the acidic region, WUS box and EAR-like motif) that may play an important role in the function of a transcription factor (Ikeda and Ohme-Takagi 2009). In the modern clade, motif 5 (WUS box with TLLFP amino acids) is present in most WOX proteins. Motif 18 (EAR-like motif with SLELSL amino acids) was detected only in WUS and WOX5 subgroups within the modern clade (Fig. S2).

Distinct expression profiles and diversification of *WOX* genes in *Salix suchowensis* tissues

Proteins control gene expression temporally and spatially, so their functions can be defined by the location and timing of the genes that encode them (Liu et al. 2014). Because the 15 *WOX* genes found in *S. suchowensis* were expressed in roots, stems and leaves, they may participate in the development of all organs. Figure 6 shows that *SsWOX1B*, *SsWOX5* and three *SsWOX13* genes were primarily expressed in the roots. *SsWOX4*, *SsWOX5*, *SsWOX9A* and three *SsWOX13* were strongly expressed in the stems, whereas *SsWOX1B*, *SsWOX4*, *SsWOX5* and three *SsWOX13* genes were highly expressed in the *S* leaves. In contrast, *SsWUSA*, *SsWUSB*, *SsWOX1A*, *SsWOX2*, *SsWOX6*, *SsWOX9B*, and *SsWOX11/12* were almost absent from roots. Additionally, we note that expression of *SsWUSA/B*, *SsWOX1A*, *SsWOX2*, *SsWOX6*, *SsWOX9B*, and *SsWOX11/12* genes were restricted to a small area or not restricted to any area (Fig. 7).

In the modern clade, no *WOX* genes from *S. suchowensis* were closely associated with *AtWOX3* or *AtWOX7*. The function of *AtWOX6* and *AtWOX7* has not been specified, but *AtWOX3* is expressed in leaves and takes part in their outgrowth (Dolzblasz et al. 2016). Furthermore, *AtWUS* can fully replace the role of *AtWOX3*, and *AtWOX4* can replace *AtWOX3* to some degree (Shimizu et al. 2009). As a consequence, in *S. suchowensis*, the function of *WOX3* genes controlling leaf development can could be achieved by other modern *SsWOX* genes. On this point, we noted that the *SsWOX1B*, *SsWOX4* and *SsWOX5* were expressed in leaves. Additionally, *SsWOX1B* was only expressed in leaves and root, to a much greater extent than *SsWOX1A*. In addition to the enrichment of *SsWOX4* expression in stems, it is weakly expressed in roots and leaves. On the other hand, *SsWOX5* is ubiquitously expressed, with a much higher level in roots than in other organs. In the intermediate clade, *SsWOX9A* is expressed more highly in both roots and stems than is *SsWOX9B*. In the ancient clade, *Arabidopsis* has three *WOX* genes, *AtWOX10*, *AtWOX13* and *AtWOX14*, as does *P. trichocarpa* and *V. vinifera* (Gambino et al. 2011; Liu et al. 2014), and *S. suchowensis* (*SsWOX13* and sister pairs *SsWOX13B* and *SsWOX13C*). They seem to represent the functional diversification of *AtWOX10*, *AtWOX13*, and *AtWOX14*. A previous study that *AtWOX13* gene functions in replum formation during fruit development (Romera-Branchat et al. 2013), and *AtWOX14* and *AtWOX4* are

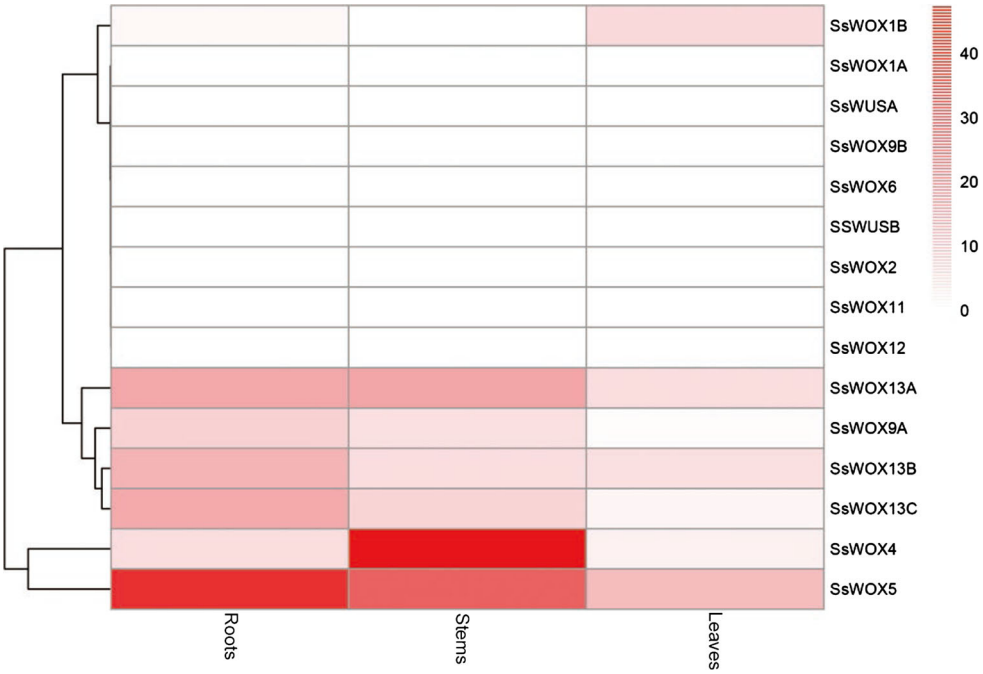

Fig. 6 Expression profiles of 15 *SsWOX* genes in roots, stems and leaves of *S. suchowensis*. Color scale represents RPKM normalized log2 transformed counts: red, highly expressed; white, not expressed

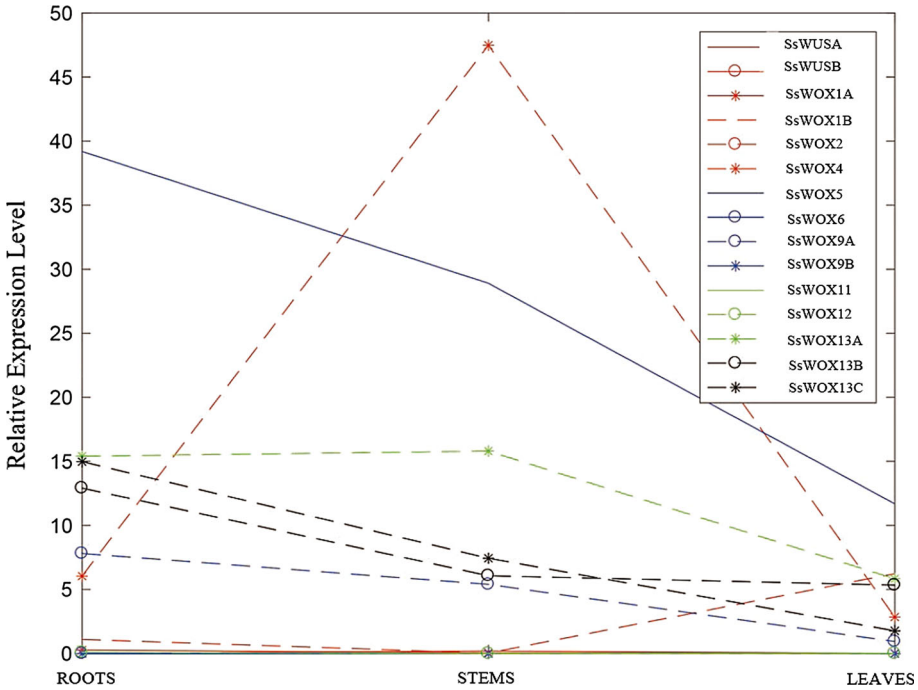

Fig. 7 Expression level of *WOX* genes in roots, stems and leaves of *S. suchowensis*

important during vascular cell division (Etchells et al. 2013). The function of *AtWOX10* is unknown. Three ancient *SsWOX* genes were ubiquitously expressed in all three organs in *S. suchowensis* (Fig. 6). *SsWOX13A* and *SsWOX13B* were expressed in roots, stems and leaves. *SsWOX13A* was expressed at much higher levels than the other three *SsWOX13* genes in roots and stems, suggesting its conserved function in root and stem development.

Disscussion

Salix suchowensis WOX family with a distinctive evolutionary pattern

Fifteen *WOX* genes of *S. suchowensis* were identified, signifying new contributions to the *WOX* family in plants. Other plants also have high numbers of *WOX* genes; e.g., together, *Arabidopsis*, *P. trichocarpa*, *O. galberrima*, *C.*

lanatus and *V. vinifera* (Graaff et al. 2009; Lian and Ding 2014) average more than 10 *WOX* genes. Whole-genome duplication (WGD) events could have contributed to the scale of the gene family. *Salix suchowensis* has *SsWOX* genes, with duplications on *SsWUS*, *SsWOX1*, *SsWOX9* and *SsWOX13*. However, 18 genes in *P. trichocarpa* are due to duplication events in the *PtrWUS*, *PtrWOX1*, *PtrWOX2*, *PtrWOX4*, *PtrWOX5* subgroups. We surmise that *WOX* genes of Salicaceae plants may regulate the same function during normal growth and development. Furthermore, 38 *WOX* genes were identified in *Gossypium hirsutum* (*G. hirsutum*), the most of any species. Polyploidization resulting from WGD and segmental duplication are important evolutionary events; the function of the new gene might be redundant and lead to evolutionary innovation (Yang et al. 2017).

Conserved sequence and diversification of WOX proteins in Salicaceae

Alignment and phylogenetic analysis of WOX protein sequences demonstrated that WOX protein sequences in Salicaceae plants have a higher similarity than in those of other eudicot plants. The high number of conserved amino acids in the homeodomain of Salicaceae sequences can also indicate their close evolutionary relationship. The presence of the WUS box in the modern clade is also consistent with previous studies (Zhang et al. 2010). Additionally, the WOXs in the ancient clade of Salicaceae are much different from those of *Arabidopsis*, suggesting a longer evolutionary history in this clade. Moreover, the absence of the WUS-box in SsWOX1 TL-LFP and EAR-like in SsWOX5 of L-L-L-S are interesting findings. The absence of a WUS box in SsWOX1A and SsWOX1B suggests that they might share the same function as AtWOX7 (Zhang et al. 2015) and require further study.

We identified 15 WOX-encoding genes in *S. suchowensis*. However, no *WOX* genes from *S. suchowensis* and *P. trichocarpa* are classified as homologs of *AtWOX3*, *AtWOX7*, *AtWOX10* or *AtWOX14*. We thus infer that function of these genes is carried on by other *WOX* genes in *S. suchowensis* and *P. trichocarpa.*

Expression of duplicated *WOX* genes in tissues

The expression patterns of the *WOX* genes in *S. suchowensis* tissues were quantified to better comprehend their functions. Some *SsWOX* genes were expressed in all three organs analyzed (roots, stems and leaves). The high expression of *SsWOX1B* and *SsWOX13* in leaves indicates a significant function in leaf development, similar to *Medicago truncatula* STENOFOLIA (Zhang et al. 2014). Compared with the expression of other *WOX* genes in the roots of *S. suchowensis*, the expression of *SsWOX13* is highest,which may indicate it has the same function as *AtWOX5* in the root apical meristem (Sarkar et al. 2007). This higher expression in the root also was found in *G. hirsutum*; the expansion of *GhWOX4* through WGD and segmental duplication indicated that these genes might have a novel function (Yang et al. 2017). In addition, *GhWOX4* responds differently when exposed to different stresses, indicating that *GhWOX4* have different functions under different situations. Furthermore, *WOX13* is a key component in the regulation of root development and expressed at a high level. Expression of *AtWOX13* and three *SsWOX13s* genes in the root and stem likely denotes their function in the development of these two organs (Zhang et al. 2010). Also in *G. hirsutum*, *GhWOX13* is expressed in almost all organs during flower and fruit development, suggesting that *WOX13* genes in the ancient clades are active during development (Yang et al. 2017). Because functional genes affect metabolite profiles, then such a study for *WOX* genes could provide a new perspectives on developmental mechanisms in *S. suchowensis* (Zhang et al. 2015). Future studies should also focus on the differential expression of *WOX* genes under various stresses.

Conclusion

The results of this study provide a genomic structure for further research on *SsWOX* genes and a new perspective on the evolution of *WOX* genes in Salicaceae plants, representing great progress in our knowledge of the functions and evolution of the 15 WOX proteins in *S. suchowensis*, especially their potential roles in root development. Our analysis of the conserved structural features and phylogenetic relationships of WOX domains suggest that chromosomal duplication events may have led to the expansion of the *SsWOX* gene, and all 15 *SsWOX* genes were assigned among the modern clade, the intermediate clade and the ancient clade. According to the expression pattern of *SsWOX* genes determined in this study, *SsWOX5* and *SsWOX13* in *S. suchowensis* play major roles in root development. WUS-box and EAR-like motifs in *S. suchowensis* WOX proteins likely fulfill additional biological functions. Evolutionary analysis of *S. suchowensis, A. thaliana* and other four eudicot plants (*P. trichocarpa*, *O. galberrima, C. lanatus* and *V. vinifera*) is helpful for understanding the role of *WOX* genes in plant development. The present annotation and analysis of the *WOX* genes and the proteins they encode in the Salicaceae family lays a foundation for breeding and genetic engineering of these and other woody plants.

References

Bailey TL, Williams N, Misleh C (2006) MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34(Web Server issue):W369

Bi C, Xu Y, Ye Q (2016) Genome-wide identification and characterization of WRKY gene family in *Salix suchowensis*. Peer J 4(9):e2437

Breuninger H et al (2008) Differential expression of WOX genes mediates apical-basal axis formation in the Arabidopsis embryo. Dev Cell 14(6):867–876

Cao Y, Han Y, Meng D (2017) Genome-wide analysis suggests the relaxed purifying selection affect the evolution of WOX genes in *Pyrus bretschneideri, Prunus persica, Prunus mume,* and *Fragaria vesca*. Front Genet 8:78

Dai X, Hu Q, Cai Q (2014) The willow genome and divergent evolution from poplar after the common genome duplication. Cell Res 24(10):1274

Dolzblasz A, Nardmann J, Clerici E (2016) Stem cell regulation by arabidopsis WOX genes. Mol Plant 9(7):1028–1039

Etchells JP, Provost CM, Mishra L (2013) WOX4 and WOX14 act downstream of the PXY receptor kinase to regulate plant vascular proliferation independently of any role in vascular organisation. Development 140(10):2224

Finn RD, Bateman A, Clements J (2016) The Pfam protein families database. Nucleic Acids Res 42(Database issue):D222–230

Gambino G, Minuto M, Boccacci P (2011) Characterization of expression dynamics of WOX homeodomain transcription factors during somatic embryogenesis in *Vitis vinifera*. J Exp Bot 62(3):1089

Ge Y, Liu J, Zeng M (2016) Identification of WOX family genes in *Selaginella kraussiana* for studies on stem cells and regeneration in lycophytes. Front Plant Sci 7(291):93

Graaff EVD, Laux T, Rensing SA (2009) The WUS homeobox-containing (WOX) protein family. Genome Biol 10(12):248

Gu Z, Cavalcanti A, Chen FC, Bouman P, Li WH (2002) Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. Mol Biol Evol 19(3):256–262

Haecker A, Grosshardt R, Geiges B (2004) Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. Development 131(3):657–668

Hirakawa Y, Kondo Y, Fukuda H (2010) TDIF peptide signaling regulates vascular stem cell proliferation via the wox4 homeobox gene in Arabidopsis. Plant Cell 22(8):2618–2629

Hu B, Jin J, Guo A (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics 31(8):1296

Ikeda M, Ohme-Takagi M (2009) Arabidopsis WUSCHEL is a bifunctional transcription factor that acts as a repressor in stem cell regulation and as an activator in floral patterning. Plant Cell 21(11):3493–3505

Ji J, Shimizu R, Sinha N, Scanlon MJ (2010a) Analyses of WOX4 transgenics provide further evidence for the evolution of the gene family during the regulation of diverse stem cell functions. Plant Signal Behav 5(7):916–920

Ji J, Strable J, Shimizu R (2010b) WOX4 promotes procambial development. Plant Physiol 152(3):1346–1356

Jin J, Tian F, Yang DC (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45(Database issue):D1040–D1045

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33(7):1870

Larkin MA, Blackshields G, Brown NP (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948

Laux T, Mayer KF, Berger J (1996) The WUSCHEL gene is required for shoot and floral meristem integrity in Arabidopsis. Development 122(1):87

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lian G, Wang Q, Ding Z (2014) Origins and evolution of WUSCHEL-related homeobox protein family in plant kingdom. Sci World J 2014(1):534140

Lin H, Niu L, McHale NA (2013) Evolutionarily conserved repressive activity of WOX proteins mediates leaf blade outgrowth and floral organ development in plants. Proc Natl Acad Sci USA 110(1):366

Liu B, Wang L, Zhang J (2014) WUSCHEL-related homeobox genes in *Populus tomentosa*: diversified expression patterns and a functional similarity in adventitious root formation. BMC Genome 15(1):296

Luan F, Wang X, Shang L (2013) A highly efficient regeneration system for watermelon (*Citrullus lanatus thunb*.). Pak J Bot 45(1):145–150

Oshchepkova EA, Omelyanchuk NA, Savina MS (2017) Systems biology analysis of the WOX5 gene and its functions in the root stem cell niche. Russ J Genet Appl Res 7(4):404–420

Romera-Branchat M, Ripoll JJ, Yanofsky MF (2013) The WOX13 homeobox gene promotes replum formation in the *Arabidopsis thaliana* fruit. Plant J 73(1):37–49

Sarkar AK, Luijten M, Miyashima S (2007) Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. Nature 446(7137):811

Shimizu R, Ji J, Kelsey E (2009) Tissue specificity and evolution of meristematic WOX3 function. Plant Physiol 149(2):841

Suer S, Agusti J, Sanchez P (2011) WOX4 imparts auxin responsiveness to cambium cells in Arabidopsis. Plant Cell 23(9):3247

Team RC (2004-2016) GUI 1.69. R: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Wang X, Cheng F, Rohlsen D (2018) Organellar genome assembly methods and comparative analysis of horticultural plants. Hortic Res 5(1):3

Yang Z, Gong Q, Qin W (2017) Genome-wide analysis of WOX genes in upland cotton and their expression pattern under different stresses. BMC Plant Biol 17(1):113

Ye N, Wang X, Li J (2017) Assembly and comparative analysis of complete mitochondiral genome sequence of an economic plant *Salix suchowensis*. Peer J 5:e3148

Zhang X, Zong J, Liu J (2010) Genome-wide analysis of WOX gene family in rice, sorghum, maize, Arabidopsis and poplar. Bull Bot 52(11):1016–1026

Zhang Y, Wu R, Qin G (2011) Over-expression of WOX1 leads to defects in meristem development and polyamine homeostasis in Arabidopsis. J Integr Plant Biol 53(6):493–506

Zhang F, Wang Y, Li G (2014) Stenofolia recruits topless to repress asymmetric LEAVES2 at the leaf margin and promote leaf blade outgrowth in *Medicago truncatula*. Plant Cell 26(2):650

Zhang N, Huang X, Bao Y (2015) Genome-wide identification and expression profiling of WUSCHEL-related homeobox (WOX) genes during adventitious shoot regeneration of watermelon (*Citrullus lanatus*). Acta Physiol Plant 37(11):1–12