

Study on the Structure and Evolution of WOX Genes in the *Populus trichocarpa* Genome

Xueqing Zhou, Zhen Xing*

Tibet Agricultural and Animal Husbandry University, Bayi, Nyingchi, Tibet, China

Abstract: The WOX gene family, a group of plant transcription factors, regulates growth, development, metabolism, and hormone signaling. This study employed bioinformatics to predict members and sequence characteristics of the *Populus trichocarpa* WOX gene family. By aligning the *Populus trichocarpa* genome with *Arabidopsis* and rice WOX genes, protein and nucleic acid sequences were obtained. Protein homology analysis identified all members, visualized with TBtools. Using MEGA11, an evolutionary tree of *Populus trichocarpa* and *Arabidopsis* WOX genes was constructed, revealing 18 predicted *Populus trichocarpa* WOX genes. Analysis of protein properties showed stability, and the evolutionary tree classified genes into branches, suggesting a close relationship between *Populus trichocarpa* and *Arabidopsis* WOX genes. Gene structure analysis revealed sequence variation and diversity.

1. Introduction

Populus trichocarpa, belonging to the genus *Populus* in the family Salicaceae, has had its gene sequencing completed in recent years. The PagWOX11/12a gene is expressed in various tissues of 84K *Populus trichocarpa* [1]. Overexpression of DcWOX4 in *Dendrobium officinale* leads to dwarfism and a two-week delay in flowering [2]. EuWOXs have low expression levels in the leaves of *Eucommia ulmoides* [3]. In *Arabidopsis*, the 15 identified WOX gene family members can be divided into three branches: the modern evolutionary branch (including WUS, AtWOX1~AtWOX7), the ancient evolutionary branch (including AtWOX10, AtWOX13, AtWOX14), and the intermediate evolutionary branch (including AtWOX8, AtWOX9, AtWOX11, AtWOX12) [4]. WOX transcription factor family members all contain an HD domain consisting of 60-66 amino acids [5]. Research on WOX genes in woody plants is limited, mostly focusing on herbaceous plants. This study explores WOX genes in *Populus trichocarpa*, analyzing its genome for gene prediction, phylogenetics, chromosome localization, and conserved motifs.

2 Materials And Methods

2.1 Prediction of WOX Genes in the *Populus trichocarpa* Genome

To identify WOX genes in the *Populus trichocarpa* genome, the full-length protein sequences of 15 *Arabidopsis* WOX members were downloaded from the TAIR public database (<http://www.arabidopsis.org/>), and

the full-length protein sequences of 14 rice WOX members were downloaded from the rice genome database. These *Arabidopsis* and rice protein sequences were used as *reference* sequences. The *Populus trichocarpa* genome sequence was downloaded from Phytozome v13. Subsequently, full-length sequence alignment was performed using MAFFT, followed by the construction of a hidden Markov model (HMM) using HMMER software. The HMM was used to search the *Populus* protein sequences (default parameters, E-value < 10⁻⁵), yielding candidate sequences. The protein structures of the candidate sequences were analyzed using Pfam (<http://pfam.xfam.org/>) and SMART (<http://smart.embl-heidelberg.de/>), and incomplete sequences were excluded, resulting in the prediction of 18 *Populus* WOX gene members. The results were validated by using the sequences obtained from the HMMER search as input for a BLAST search (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) against the *Populus* genome database, yielding 18 sequences. Additionally, *Arabidopsis* sequences were used for a BLAST search to obtain sequences. Finally, a custom script was used to annotate the two sets of predicted *Populus* WOX gene sequences using the InterProScan database (<http://www.ebi.ac.uk/interpro/search/sequence/>), and duplicate and blank sequences were manually removed. Subsequently, the number of amino acids, molecular weight, isoelectric point, protein hydrophobicity, and aliphatic index of the *Populus* WOX proteins were calculated using the ExPASy online tools (<http://www.expasy.org/tools/>). The subcellular localization of the *Populus* WOX protein family members was predicted using the CELLO online tool (<http://cello.life.nctu.edu.tw/>). The secondary structure of the proteins was predicted using the SOPMA online tool

*Corresponding author: xztibetan@163.com

(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html).

2.2 Phylogenetic and Gene Structure of WOX Genes in the *Populus trichocarpa* Genome

The *Populus* WOX protein sequences, along with *Arabidopsis* and rice WOX protein sequences, were aligned using MUSCLE in the MEGA11 software. False-positive genes were removed, and a phylogenetic tree was constructed using the neighbor-joining method (NJ) with default parameters and a bootstrap value of 1000. The tree was saved in nwk format and then visualized using the ITOL website. The *Populus* WOX genome sequences and CDS sequences were uploaded to GSDS 2.0 (<http://gsds.cbi.pku.edu.cn>) for gene structure analysis. The *Populus* WOX protein sequences were uploaded to the MEME online tool, with the number of motifs set to 15, and the results were visualized using the TBtools software.

2.3 Conserved Motifs of WOX Genes in the *Populus trichocarpa* Genome

Submit the protein file and gene IDs of *Populus trichocarpa* WOX genes to TBtools (<https://github.com/CJ-Chen/TBtools>) and save them in FASTA format. Then, open MEME (<http://meme-suite.org>), upload the FASTA file, choose 15 motifs, and save the results. Next, construct a phylogenetic tree containing only the protein sequences of *Populus trichocarpa* in MEGA11. Then, analyze the results in TBtools software to generate a figure for the analysis of conserved motifs of genes.

3 Results And Analysis

3.1 Prediction Analysis of WOX Genes in the *Populus trichocarpa* Genome

In this study, we employed a combination of Pfam (<http://pfam.xfam.org/>), Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), and Interproscan (<http://www.ebi.ac.uk/interpro/search/sequence/>) to predict and cross-validate each other. Sequences lacking the WOX domain were removed. A total of 18 WOX gene family members were predicted from the entire genome of *Populus trichocarpa*.

3.2 Analysis of Physicochemical Properties and Protein Structure Prediction

Analysis of Protein Structure Prediction for the *Populus trichocarpa* WOX Family Members (Table 1): The results indicate that the secondary structure prediction of the *Populus* WOX gene family proteins reveals that all 18 proteins contain α -helices, β -turns, random coils, and extended chains. α -helices and random coils are predominant, followed by β -turns and extended chains, suggesting that the *Populus* WOX proteins are mainly composed of α -helices and random coils. Subcellular prediction results show that 17 out of the 18 *Populus* WOX gene proteins are localized in the nucleus, indicating that these proteins function predominantly in the nucleus.

Analysis of Physicochemical Properties of *Populus trichocarpa* WOX Family Members (Table 2): The results show significant differences in the number of amino acids in the *Populus* WOX proteins, ranging from 181aa to 387aa. The molecular weight ranges from 21012.58 to 43971.74 kDa. The theoretical isoelectric points (PI) range from 5.33 to 9.15, with 9 acidic and 9 basic proteins, indicating an equal distribution of acidic and basic amino acids in the *Populus* WOX proteins. The hydrophobicity values range from -0.452 to -1.089, indicating that these proteins are hydrophilic. The aliphatic index ranges from 45.08 to 71.35. These results demonstrate that different members of the *Populus* WOX gene family exhibit diverse physical and chemical properties. It can be inferred that different members of the *Populus* WOX gene family may have distinct biological functions.

Table 1. Protein secondary structure prediction of *Populus trichocarpa* wox genes

Protein name	α -helix/%	β -turns/%	Random coil	Extended strand
Potri.014G025300.1	27.23	3.76	60.09	8.92
Potri.007G012100.1	16.29	5.3	64.02	14.39
Potri.009G029200.1	21.58	4.15	65.98	8.3
Potri.019G040800.2	13.31	6.05	66.13	14.52
Potri.015G039100.1	28.34	2.67	58.82	10.16
Potri.002G124100.1	30.52	3.76	59.62	6.10
Potri.002G008800.1	33.95	2.79	59.07	4.17
Potri.011G061400.1	12.47	3.98	72.15	11.41
Potri.012G047700.1	26.36	3.1	59.95	10.59
Potri.001G237900.1	18.37	4.08	67.76	9.8
Potri.004G051600.3	13.54	4.17	69.27	13.02
Potri.010G111400.1	27.85	1.9	59.81	10.44

Potri.010G192100.1	36.46	5.52	45.3	12.71
Potri.005G114700.1	20.83	3.79	65.15	10.23
Potri.005G101800.1	33.47	3.63	55.24	7.66
Potri.005G252800.1	39.35	6.94	49.07	4.63
Potri.008G065400.1	38.67	8.84	41.99	10.5
Potri.013G066900.1	13.78	7.09	65.35	13.78

Table 2. Physicochemical properties of WOX protein in *Populus trichocarpa*

Genome entry Number	Number of amino acids(aa)	The molecular Weight(KD)	Theoretical isoelectric point	Protein hydrophobicity	Aliphatic amino acid index	Prediction of subcellular localization
Potri.014G025300.1.p	213	24467.54	9.03	- 1.027	57.70	nucleus
Potri.007G012100.1.p	264	29279.84	6.43	- 0.917	49.17	nucleus
Potri.009G029200.1.p	241	27042.31	8.22	- 0.691	50.62	nucleus
Potri.019G040800.2.p	248	27466.51	6.16	- 0.452	66.69	nucleus
Potri.015G039100.1.p	374	42333.79	6.61	- 0.875	55.11	nucleus
Potri.002G124100.1.p	213	24421.46	9.15	- 0.991	58.64	nucleus
Potri.002G008800.1.p	215	24483.58	5.33	- 0.775	61.72	mitochondrion
Potri.011G061400.1.p	377	41608.59	7.75	- 0.486	71.35	nucleus
Potri.012G047700.1.p	387	43971.74	5.62	- 0.783	57.52	nucleus
Potri.001G237900.1.p	245	27898.20	8.34	-0.757	58.94	nucleus
>Potri.004G051600.3.p	384	42411.53	8.78	-0.537	61.41	nucleus
Potri.010G111400.1.p	316	35705.69	8.31	-0.885	55.57	nucleus
Potri.010G192100.1.p	181	20891.38	8.36	-0.784	55.52	nucleus
Potri.005G114700.1.p	264	30043.61	6.33	-1.089	45.08	nucleus
Potri.005G101800.1.p	248	27988.99	5.51	-0.887	66.85	nucleus
Potri.005G252800.1.p	216	24642.89	5.73	-0.756	68.15	nucleus
Potri.008G065400.1.p	181	21012.58	8.99	-0.810	58.73	nucleus
Potri.013G066900.1.p	254	28081.10	5.87	-0.511	65.55	nucleus

3.3 Phylogenetic Analysis of WOX Genes in the *Populus trichocarpa* Genome

To explore the evolutionary relationship among *Populus trichocarpa*, *Arabidopsis*, and rice WOX members, the full-length sequences were aligned using MUSCLE in MEGA11 software, and a phylogenetic tree was constructed as shown in Figure 1. According to the classification in *Arabidopsis*, these WOX members can be divided into three clades: ancient clade, intermediate clade, and modern clade (WUS clade). The WUS clade is the largest branch on the tree and represents the largest clade in *Arabidopsis*, rice, and *Populus*. Within the WUS clade, there are 11 WOX members in *Populus trichocarpa*, 3 in the ancient clade, and 4 in the intermediate clade. The results indicate that WOX gene members on the same branch are closely related, while those on different branches are more distantly related. Additionally, there may be a close evolutionary relationship between *Populus trichocarpa* and *Arabidopsis* WOX members. Based on this, it can be inferred that the WOX gene family in *Populus trichocarpa* has undergone some variations during its evolutionary history, leading to WOX genes with different functions in plant development.

3.4 Analysis of WOX Gene Structure and Conserved Motifs in the *Populus trichocarpa* Genome

Analysis of gene structure is instrumental in revealing the evolutionary patterns of gene families. As depicted in Figure 2, the results of gene structure analysis show that all family members contain both exons and introns. The number of introns in *Populus trichocarpa* WOX gene family members ranges from 1 to 4, while the number of exons ranges from 2 to 4. On the same branch, the lengths and positions of exons and introns vary, but their orders remain the same. Significant differences in exons and introns are observed among different branches. The gene structure of *Populus trichocarpa* WOX gene family members is relatively similar and conservative. Based on these findings, it can be inferred that the diversity of gene functions in organisms may be attributed to the decrease or increase in introns during the evolutionary history of the gene family.

The results of the conserved motif analysis of proteins show that among the 18 WOX gene proteins in *Populus trichocarpa*, 15 contain conserved motifs, designated as Motif 1 to 15. Motif 1 and 2 are identified as core conserved sequences. Across different branches, the number of motifs ranges from 3 to 8, with similar positions. Motifs within the evolutionary branches exhibit similar positions, lengths, and sequences, suggesting potential functional similarities. Significant differences in

conserved motifs between different evolutionary branches may be associated with the functions of *Populus*

trichocarpa WOX genes, consistent with the phylogenetic relationships depicted in Figure 3.

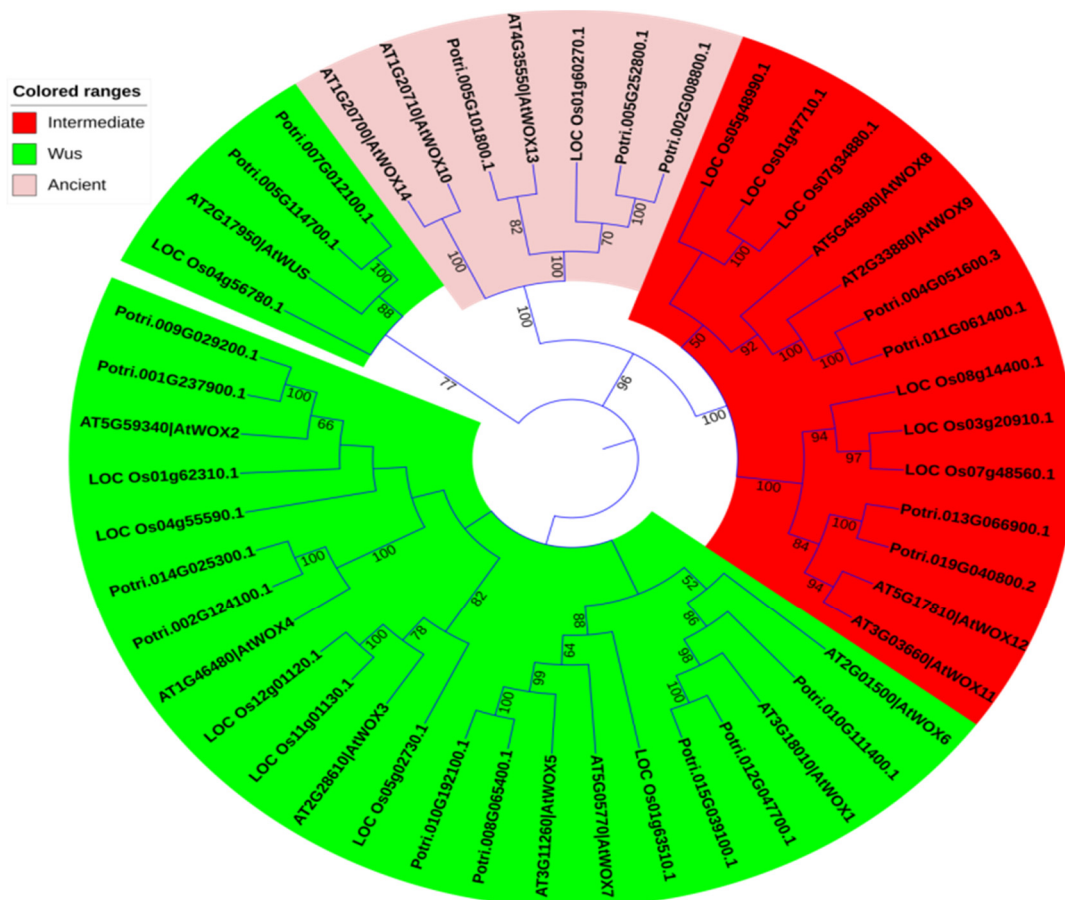


Fig. 1. Phylogenetic tree analysis of WOX gene family in *Populus trichocarpa*

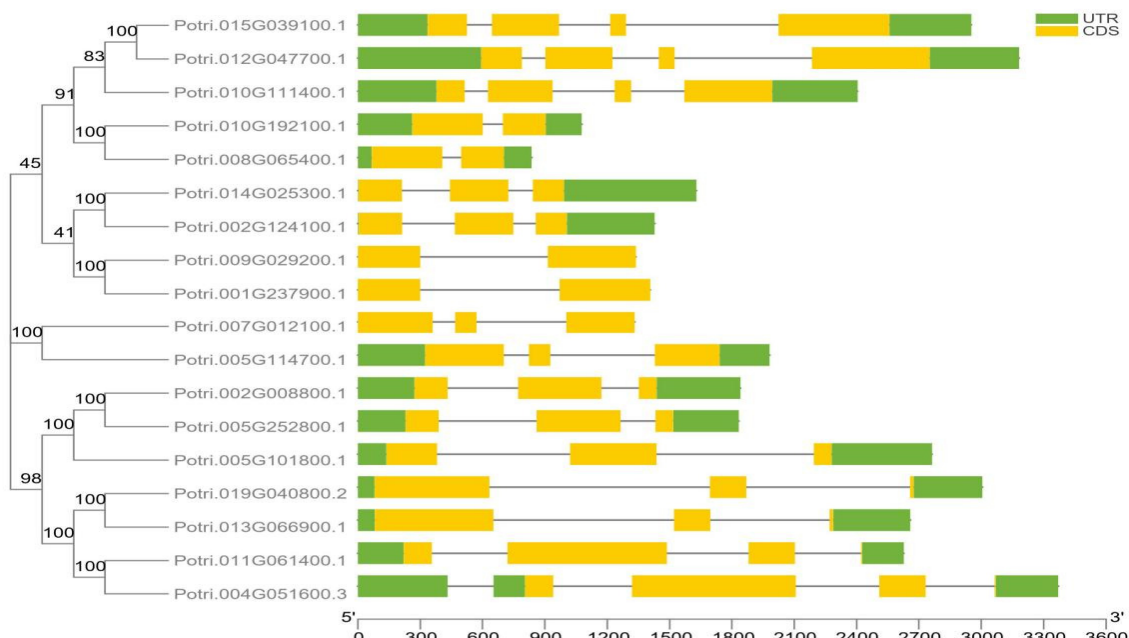


Fig. 2. Gene structure analysis of WOX gene family in *Populus trichocarpa*



Fig. 3. Protein motifs of *Populus* WOX transcription factors

4 Conclusion

In this study, the *Populus trichocarpa* WOX gene family was predicted using Pfam, Blast, and Interproscan programs, resulting in the identification of 18 members, speculated to be the result of gene duplication events. Analysis revealed that WOX proteins are predominantly located in the cell nucleus, and the phylogenetic tree showed that the 18 genes are distributed among three branches: WUS, ancient, and intermediate, indicating close evolutionary relationships with the *Arabidopsis* WOX family. Structural analysis unveiled that all members contain introns and exons, with significant variations in their numbers. Conserved motif analysis demonstrated that Motif 1 and 2 represent core sequences, exhibiting high consistency within the same evolutionary branch, reflecting both evolutionary diversity and conservation. This study provides a theoretical basis for further research on *Populus trichocarpa* WOX genes.

Acknowledgments

This work was financially supported by ‘Study on the support system of modern forestry technology in Tibet (2018xz503118002)’; ‘Study on the model of planting and breeding circular agriculture in Tibet’;

References

1. Wen S S, Wang L Q, Lu M Z. The impact of the PagWOX11/12a gene on stem growth and development in *Populus alba* × *Populus glandulosa* [J]. Forest Research, 2023, 36(01): 39-46.
2. Chen K, Wang H, Chen Y T, et al. Functional analysis of WOX gene family in the growth and development of *Dendrobium officinale* [J]. Hereditas, 2023, 45(08): 700-714. DOI:10.16288/j.ycz.22-392.
3. Liu J, Li L, Chen Y L, et al. Identification and expression of WOX gene family in leaf development of *Eucommia ulmoides* [J]. Journal of Zhejiang A&F University, 2023, 40(01): 1-11.
4. van der Graaff E, Laux T, Rensing SA. The WUS homeobox-containing (WOX) protein family. Genome Biol, 2009, 10(12): 248.
5. Gehring W J, et al. Homeodomain-DNA Recognition. Cell. 1994, 78(2): 211-223.