WUSCEL-related homeobox gene family phylogeny in crop species

Minjeong Kang[1]

[1]Interdepartmental Plant biology, Iowa State University, Ames, IA, USA

1. Introduction

The WUSCHEL-RELATED HOMEOBOX (WOX) gene family is one of the classes of plant homeobox proteins which involve in the important developmental processes, such as stem-cell maintenance and organ formation in shoot apical meristem (Alvarez et al., 2018). The WOX genes contain conserved homeodomain, which has 60 amino acid residues with the helix-loop-helix-turn-helix structure (Zhang et al., 2010). These WOX genes play important roles in *Arabidopsis* organizing center (OC) through the regulatory loop between CLAVATA (CLV) and WUSCHEL that involves the maintenance of the shoot apical meristem (Schoof et al., 2000).

The WOX protein family members can regulate cell differentiation of plant cell as both of activator and repressor (Zuo et al., 2002). The overexpressed maize WUS2 gene can dramatically improve the shoot regeneration of sorghum, rice, and sugarcane (Lowe et al., 2016). Also, none of the maize WUS orthologous expresses in *Arabidopsis*, and the CLV orthologue – WUS loop of grass differed from dicot plants (Nardmann and Werr, 2006). The previous research indicates that divergence of WOX genes among plant species may be related to the various differentiation pattern of the plant cell.

Phylogenetic analysis is the way to investigate the evolutionary development of organisms. Maximum likelihood (ML) is optimality criterion to select evolutionary trees. The likelihood (*L*) is defined as the probability of the data (*D*) given a model of evolution (*M*) and other parameters (*p*), such as model parameters and a tree topology (Posada et al., 2004).

$$L = P\,(\,D\,|\,M,\,p\,)$$

The ML requires at least approximately correct model assumptions, such as GTR, the generalized time-reversible model (Tavaré, 1896), to derive a result. When the model is not enough to describe the evolution of the sequence, the ML may not guarantee good results (Allman and Rhodes, 2005).

In the present study, a total 46 WOX members from a turnip, tomato, potato, rice, maize, sorghum, *Arabidopsis* and green algae were analyzed by maximum likelihood method to investigate the divergence of WOX genes in modern crops.

2. Materials and Methods

Search of WOX proteins and domain identification

The sequence used in this paper were obtained from EnsemblPlants (https://plants.ensembl.org) and NCBI (https://www.ncbi.nlm.nih.gov). Total 44 WOX protein sequences from dicots: *Arabidopsis thaliana, Brassica rapa, Solanum lycopersicum,* and *Solanum tuberosum,* monocots: *Oryza sativa* (japonica group)*, Zea mays, and Sorghum bicolor* and green algae: *Ostreococcus lucimarinus* and *Ostreococcus tauri*. The WOX names used in this paper is based on the description in database and isoforms are marked as a and b (Supplementary Table 1). In order to examine the homeodomain of sequences, the motif analysis was carried out using MOTIF search (http://www.genome.jp/tools/motif/) and MAST (version 4.12.0) (Bailey and Gribskov, 1998).

Multiple sequence alignment and Phylogenetic tree construction

Multiple sequence alignments were conducted by using MAFFT (version 7.245) with auto strategy setting. Maximum likelihood analysis was conducted with the following parameters by RAxML (version 8.2.9): PROTGAMMAAUTO and bootstrap (100 replicates). The Dendroscope software (version 3.5.9) was used to visualize the tree.

3. Results

Analysis of conserved domains

The motif discovery was conducted to identify the conserved homeodomain in WOX family proteins. All protein sequences used for this study showed the characteristic homeodomain of WOX genes which is 60 aa helix-loop-helix-turn-helix structure (Figure 1 and Supplementary Table 2) (Lian et al., 2014).
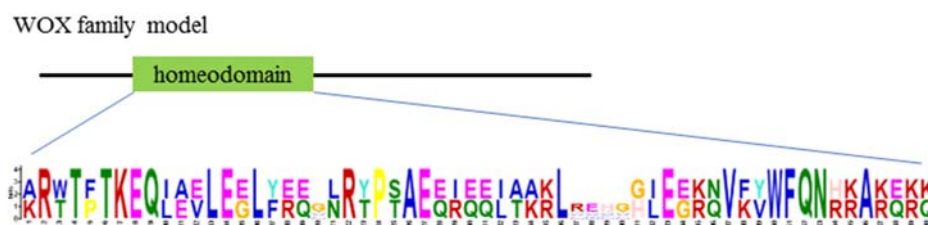

Figure 1. Conserved residues of WOX homeodomain sequence. The image was obtained by MAST (version 4.12.0).

Phylogenetic analysis of WOXs

Phylogenetic analysis was conducted to investigate the divergence of WOXs among crops. Figure 2 shows the ML tree with the bootstrap values. Phylogenetic tree of the WOX proteins have been divided into three clades: the WUS clade which specific to seed plants and ferns, the ancient clade which represents the earliest diverging WOX genes, and the intermediate clade which is interspersed between two clades (Graaff et al., 2009). Corresponded with the previous studies (Graaff et al., 2009; Lian et al., 2014), the WOXs were divided into three groups, designated WOX 1 - 7 and WUS in WUS clade, WOX 8, 9, 11, and 12 in intermediate clade, and WOX 10, 13 and 14 in ancient clade with high bootstrap value (Figure 2 and 3). However, the bootstrap values in interior edges of WUS clade is relatively lower than other clades. This result suggests that WUS clade contains more divergence members compared to other clades. Moreover, unlike other taxa, the OsWox7 and OsWox10 were found in the intermediate clade. Generally, each monocot and dicots formed subgroups in the clades. This result suggests that WOX gene divergence occurred before the speciation of monocot and eudicot.
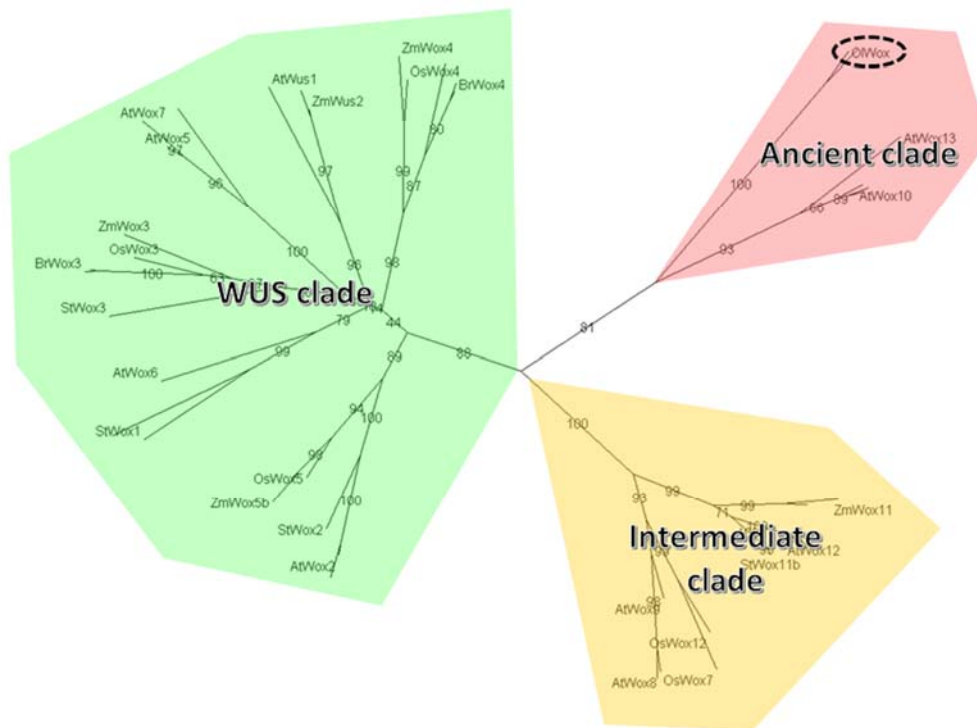

Figure 2. Phylogenic tree of WOX proteins. The tree contains three clades; WUS clade, ancient clade and intermediate clade. The black dotted circle indicates the green algae.

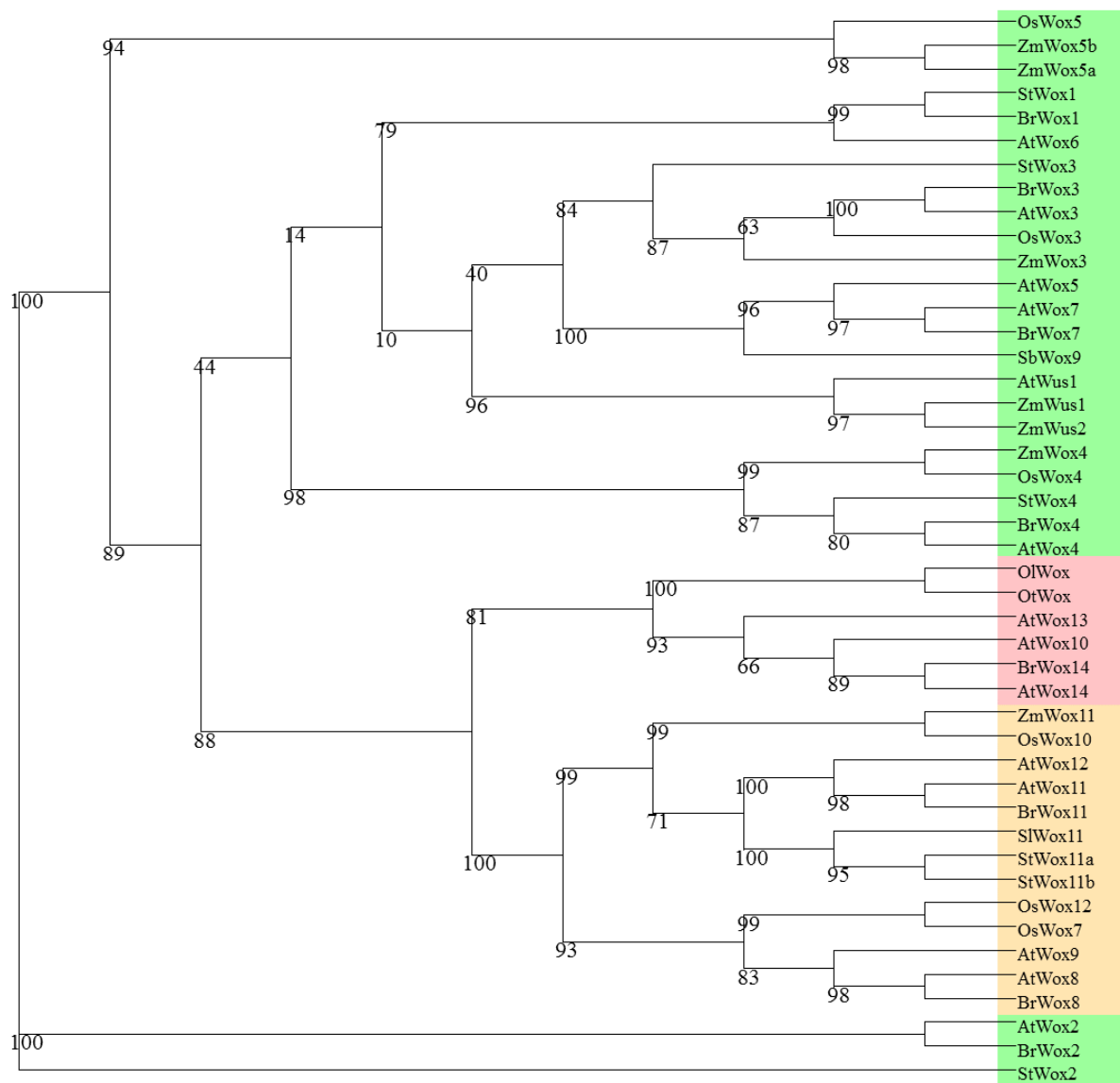**Figure 3.** Unrooted phylogenetic tree of 46 WOX proteins with bootstrap value. The box color represents the clade of taxa. The green is the WUS clade, the red is ancient clade, and the orange is intermediate clade.

4. Discussion

In this study, most of WOX proteins formed clades with its homologs proteins and WOXs from green algae were included to the ancient clade. However, the OsWox7 and OsWox10 showed ectopic location. Previous research indicates that the WOXs from monocots form 3-5 clades unlike AtWOXs and maize WOX2 protein expression pattern is different from the WOX2 homologous protein of *Arabidopsis* (Nardmann and Werr, 2006; Nardmann et al., 2007; Lian et al., 2014). Therefore, it is possible that the differentiation of several WOXs may be occurred after speciation of monocot and dicot. To investigate its functional divergences, 3D structures of WOX protein is needed for the future study.

The crops important in agriculture field were analyzed to investigate the various differentiation pattern among the species. The result of the phylogenetic analysis is insufficient to explain the cell development patterns. It suggests that other factors such as cis-regulatory region or methylation may involve in the expression and function of WOX gene family.

References

Allman, E.S., and J.A. Rhodes. 2005. Lecture Notes: The Mathematics of Phylogenetics. : 126p.

Alvarez, J.M., N. Bueno, R.A. Cañas, C. Avila, F.M. Cánovas, and R.J. Ordás. 2018. Analysis of the WUSCHEL-RELATED HOMEOBOX gene family in Pinus pinaster: New insights into the gene family evolution. Plant Physiol. Biochem. 123(September 2017): 304–318.

Bailey, L.T., and M. Gribskov. 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14(1): 48–54.

van der Graaff, E., T. Laux, and S.A. Rensing. 2009. The WUS homeobox-containing (WOX) protein family. Genome Biol. 10(12): 248.

Lian, G., Z. Ding, Q. Wang, D. Zhang, and J. Xu. 2014. Origins and evolution of WUSCHEL-related homeobox protein family in plant kingdom. Sci. World J. 2014(2009): 1–13.

Lowe, K., E. Wu, N. Wang, G. Hoerster, C. Hastings, M.-J. Cho, C. Scelonge, B. Lenderts, M. Chamberlin, J. Cushatt, L. Wang, L. Ryan, T. Khan, J. Chow-Yiu, W. Hua, M. Yu, J. Banh, Z. Bao, K. Brink, E. Igo, B. Rudrappa, P. Shamseer, W. Bruce, L. Newman, B. Shen, P. Zheng, D. Bidney, C. Falco, J. Register, Z.-Y. Zhao, D. Xu, T. Jones, and W. Gordon-Kamm. 2016. Morphogenic Regulators *Baby boom* and *Wuschel* Improve Monocot Transformation. Plant Cell 28(9): 1998–2015.

Nardmann, J., and W. Werr. 2006. The Shoot Stem Cell Niche in Angiosperms: Expression Patterns of WUS Orthologues in Rice and Maize Imply Major Modifications in the Course of Mono- and Dicot Evolution. Mol. Biol. Evol. 23(12): 2492–2504.

Nardmann, J., R. Zimmermann, D. Durantini, E. Kranz, and W. Werr. 2007. WOX gene phylogeny in poaceae: A comparative approach addressing leaf and embryo development. Mol. Biol. Evol. 24(11): 2474–2484.

Posada, D., T.R. Buckley, and J. Thorne. 2004. Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests (J Thorne, Ed.). Syst. Biol. 53(5): 793–808.

Schoof, H., M. Lenhard, A. Haecker, K.F.. Mayer, G. Jürgens, and T. Laux. 2000. The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. Cell 100(6): 635–644.

Tavaré, S. 1896. Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Am. Math. Soc. 17: 57–86.

Zhang, X., J. Zong, J. Liu, J. Yin, and D. Zhang. 2010. Genome-wide analysis of WOX gene family in rice, sorghum, maize, arabidopsis and poplar. J. Integr. Plant Biol. 52(11): 1016–1026.

Zuo, J., Q.-W. Niu, G. Frugis, and N.-H. Chua. 2002. The WUSCHEL gene promotes vegetative-to-embryonic transition in Arabidopsis. Plant J. 30(3): 349–359.

Supplementary Table 1. Accession numbers for all protein used in phylogenetic analysis

Name	Accession number	Source
AtWus1	NP_565429	https://www.ncbi.nlm.nih.gov/
AtWox1	AAP37133	https://www.ncbi.nlm.nih.gov/
AtWox2	AAP37131	https://www.ncbi.nlm.nih.gov/
AtWox3	AAP37135	https://www.ncbi.nlm.nih.gov/
AtWox4	AAP37134	https://www.ncbi.nlm.nih.gov/
AtWox5	AAP37136	https://www.ncbi.nlm.nih.gov/
AtWox6	AAP37137	https://www.ncbi.nlm.nih.gov/
AtWox7	NP_196196	https://www.ncbi.nlm.nih.gov/
AtWox8	AAP37138	https://www.ncbi.nlm.nih.gov/
AtWox9	AAP37139	https://www.ncbi.nlm.nih.gov/
AtWox10	NP_173494	https://www.ncbi.nlm.nih.gov/
AtWox11	NP_001118563	https://www.ncbi.nlm.nih.gov/
AtWox12	AAP37141	https://www.ncbi.nlm.nih.gov/
AtWox13	AAP37142	https://www.ncbi.nlm.nih.gov/
AtWox14	NP_173493	https://www.ncbi.nlm.nih.gov/
BrWox1	XP_009135563.1	https://www.ncbi.nlm.nih.gov/
BrWox2	XP_009120441.1	https://www.ncbi.nlm.nih.gov/
BrWox3	XP_009140922.1	https://www.ncbi.nlm.nih.gov/
BrWox4	XP_009145045.1	https://www.ncbi.nlm.nih.gov/
BrWox7	XP_009122174.1	https://www.ncbi.nlm.nih.gov/
BrWox8	XP_009128835.1	https://www.ncbi.nlm.nih.gov/
BrWox11	XP_009130381.1	https://www.ncbi.nlm.nih.gov/
BrWox14	XP_009149550.1	https://www.ncbi.nlm.nih.gov/
OsWox3	XM_015764119.1	https://www.ncbi.nlm.nih.gov/
OsWox4	XP_015635367.1	https://www.ncbi.nlm.nih.gov/
OsWox5	Q8LR86	http://www.uniprot.org/uniprot/Q8LR86
OsWox7	Os01t0667400-01	http://rapdb.dna.affrc.go.jp/
OsWox10	XM_015794156.1	https://www.ncbi.nlm.nih.gov/
OsWox12	Os05t0564500-00	http://rapdb.dna.affrc.go.jp/
SbWox9	XM_002458736	https://www.ncbi.nlm.nih.gov/
SlWox11	XP_010322493.1	https://www.ncbi.nlm.nih.gov/
StWox1	XP_006341531.1	https://www.ncbi.nlm.nih.gov/
StWox2	XP_006350912.1	https://www.ncbi.nlm.nih.gov/
StWox3	PGSC0003DMG400025358	http://solanaceae.plantbiology.msu.edu/
StWox4	XP_006354857.1	https://www.ncbi.nlm.nih.gov/
StWox11a	XP_015164268.1	https://www.ncbi.nlm.nih.gov/
StWox11b	XP_006347326.1	https://www.ncbi.nlm.nih.gov/
ZmWus1	CAJ84136	https://www.ncbi.nlm.nih.gov/
ZmWus2	NP_001105961	https://www.ncbi.nlm.nih.gov/
ZmWox3	CAM32346	https://www.ncbi.nlm.nih.gov/
ZmWox4	CAM32347	https://www.ncbi.nlm.nih.gov/
ZmWox5a	CAM32348	https://www.ncbi.nlm.nih.gov/
ZmWox5b	NP_001306687	https://www.ncbi.nlm.nih.gov/
ZmWox11	NP_001147238	https://www.ncbi.nlm.nih.gov/
OlWox	OSTLU_27102	http://www.uniprot.org/uniprot/A4S6F9
OtWox	XP_003082606.1	https://www.ncbi.nlm.nih.gov/

Supplementary Table 2. Result of Motifinder

Query	Pfam	Description
AtWox1	Homeobox	PF00046, Homeobox domain
AtWox2	Homeobox	PF00046, Homeobox domain
AtWox3	Homeobox	PF00046, Homeobox domain
AtWox4	Homeobox	PF00046, Homeobox domain
AtWox5	Homeobox	PF00046, Homeobox domain
AtWox6	Homeobox	PF00046, Homeobox domain
AtWox7	Homeobox	PF00046, Homeobox domain
AtWox8	Homeobox	PF00046, Homeobox domain
AtWox9	Homeobox	PF00046, Homeobox domain
AtWox10	Homeobox	PF00046, Homeobox domain
AtWox11	Homeobox	PF00046, Homeobox domain
AtWox12	Homeobox	PF00046, Homeobox domain
AtWox13	Homeobox	PF00046, Homeobox domain
AtWox14	Homeobox	PF00046, Homeobox domain
AtWus1	Homeobox	PF00046, Homeobox domain
ZmWus1	Homeobox	PF00046, Homeobox domain
ZmWus2	Homeobox	PF00046, Homeobox domain
ZmWox5a	Homeobox	PF00046, Homeobox domain
ZmWox5b	Homeobox	PF00046, Homeobox domain
ZmWox3	Homeobox	PF00046, Homeobox domain
ZmWox4	Homeobox	PF00046, Homeobox domain
ZmWox11	Homeobox	PF00046, Homeobox domain
SbWox9	Homeobox	PF00046, Homeobox domain
SlWox11	Homeobox	PF00046, Homeobox domain
StWox4	Homeobox	PF00046, Homeobox domain
StWox2	Homeobox	PF00046, Homeobox domain
StWox1	Homeobox	PF00046, Homeobox domain
StWox11a	Homeobox	PF00046, Homeobox domain
StWox11b	Homeobox	PF00046, Homeobox domain
StWox3	Homeobox	PF00046, Homeobox domain
OsWox7	Homeobox	PF00046, Homeobox domain
OsWox12	Homeobox	PF00046, Homeobox domain
OsWox4	Homeobox	PF00046, Homeobox domain
OsWox5	Homeobox	PF00046, Homeobox domain
OsWox10	Homeobox	PF00046, Homeobox domain
OsWox3	Homeobox	PF00046, Homeobox domain
BrWox8	Homeobox	PF00046, Homeobox domain
BrWox14	Homeobox	PF00046, Homeobox domain
BrWox2	Homeobox	PF00046, Homeobox domain
BrWox7	Homeobox	PF00046, Homeobox domain
BrWox11	Homeobox	PF00046, Homeobox domain
BrWox4	Homeobox	PF00046, Homeobox domain
BrWox1	Homeobox	PF00046, Homeobox domain
BrWox3	Homeobox	PF00046, Homeobox domain
OlWox	Homeobox	PF00046, Homeobox domain
OtWox	Homeobox	PF00046, Homeobox domain