WUSCEL-related homeobox gene family phylogeny in crop species

Minjeong Kang

1. Introduction

The cell fate during plant cell development is specified by various genes and its regulation. The WUSCHEL-RELATED HOMEOBOX (WOX) gene family 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 play important role in *Arabidopsis* organizing center (OC). A regulatory loop between CLAVATA (CLV) and WUSCHEL involves the maintenance of the shoot apical meristem (Schoof et al., 2000). The WOX gene family is a subgroup of homeodomain (HD), containing transcription factors, which has 60 amino acid residues with the helix-loop-helix-turn-helix structure (Zhang et al., 2010). Phylogenetic tree of the WOX proteins is spontaneously divided into three clades: the WUS clade, the ancient clade which represents the earliest diverging WOX genes, and the intermediate clade which is interspersed between two clades (Graaff et al., 2009).

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).

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

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, 44 WOX members of turnip, tomato, potato, rice, maize, sorghum and *Arabidopsis* were analyzed by maximum likelihood. The phylogeny tree is grouped into three clades and two ectopic OsWOXs were identified.

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,* and monocots: *Oryza sativa* (japonica group)*, Zea mays, and Sorghum bicolor*. The WOX names used in this paper is based on the description in database and isoforms are marked as a and b (see Supplementary Table 1). In order to examine the 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.
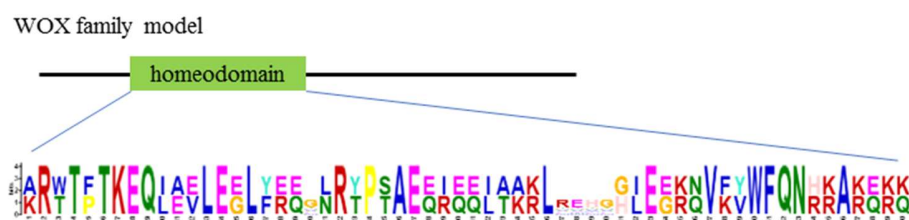
3. Results

Analysis of conserved domains

The motif discovery of 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).

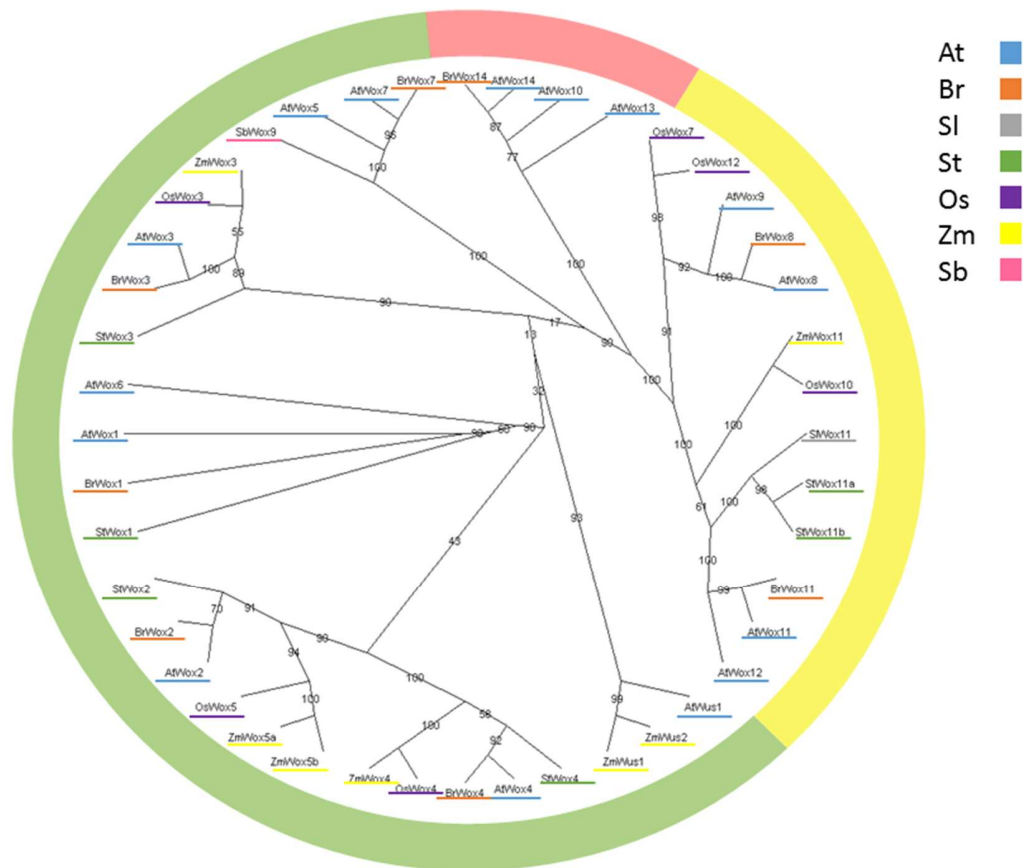*Phylogenetic analysis of WOXs*

Phylogenetic analysis was conducted to investigate the divergence of WOXs among crops. Figure 2 shows the ML tree with the bootstrap values. Corresponded with the previous studies (Graaff et al., 2009; Lian et al., 2014), the WOXs were divided into three groups, designated WOX1 – WOX7 and WUS in WUS clade, WOX 8, 9, 11, and 12 in intermediate clade, and WOX 10, 13 and 14 in ancient clade except OsWox7 and OsWox10 (Figure 2 and 3). The bootstrap values in interior edges of WUS clade is relatively lower than other clades. 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.

4. Discussion

Although most of WOX proteins form clades with its homologs proteins, the OsWox7 and OsWox10 showed ectopic location. Previous research indicates that the WOXs from monocots form 3-5 clades unlike AtWOXs (Nardmann and Werr, 2006). Therefore, it is possible that the differentiation of several WOXs may occurred after speciation of monocot and eudicot.

Figure 1. Conserved residues of WOX homeodomain sequence

Figure 2. The radial cladogram tree. The color underline represents each crop specie. Outer circle indicates the clade of taxa. The green is the WUS clade, the red is ancient clade, and the yellow is intermediate clade.

References

Allman, E.S., and J.A. Rhodes. 2005. Lecture Notes: The Mathematics of Phylogenetics. : 126 Available at https://jarhodesuaf.github.io/PhyloBook.pdf (verified 12 April 2018).

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.

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.

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.

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.

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

Name	Accession number
AtWus1	NP_565429
AtWox1	AAP37133
AtWox2	AAP37131
AtWox3	AAP37135
AtWox4	AAP37134
AtWox5	AAP37136
AtWox6	AAP37137
AtWox7	NP_196196
AtWox8	AAP37138
AtWox9	AAP37139
AtWox10	NP_173494
AtWox11	NP_001118563
AtWox12	AAP37141
AtWox13	AAP37142
AtWox14	NP_173493
BrWox1	XP_009135563.1
BrWox2	XP_009120441.1
BrWox3	XP_009140922.1
BrWox4	XP_009145045.1
BrWox7	XP_009122174.1
BrWox8	XP_009128835.1
BrWox11	XP_009130381.1
BrWox14	XP_009149550.1
OsWox3	XM_015764119.1
OsWox4	XP_015635367.1
OsWox5	Q8LR86
OsWox7	Os01t0667400-01
OsWox10	XM_015794156.1
OsWox12	Os05t0564500-00
SbWox9	XM_002458736
SlWox11	XP_010322493.1
StWox1	XP_006341531.1
StWox2	XP_006350912.1
StWox3	PGSC0003DMG400025358
StWox4	XP_006354857.1
StWox11a	XP_015164268.1
StWox11b	XP_006347326.1
ZmWus1	CAJ84136
ZmWus2	NP_001105961
ZmWox3	CAM32346
ZmWox4	CAM32347
ZmWox5a	CAM32348
ZmWox5b	NP_001306687
ZmWox11	NP_001147238

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