Scientific Research Publishing

Genome-Wide Identification, Classification and Evolutionary Expansion of KNOX Gene Family in Rice (*Oryza sativa*) and *Populus* (*Populustrichocarpa*)

Haizheng Xiong[1,2], Ainong Shi[2], Dianxing Wu[1], Yuejin Weng[2], Jun Qin[2], Waltram Second Ravelombola[2], Xiaoli Shu[1*], Wei Zhou[2*]

[1]State Key Lab of Rice Biology, Institute of Nuclear Agricultural Sciences, Zhejiang University, Hangzhou, China
[2]University of Arkansas, Division of Agriculture, Dept. of Horticulture 316 PTSC, University of Arkansas, Fayetteville, AR, USA
Email: *shuxl@zju.edu.cn, *wzhou.yan@gmail.com

Abstract

The KNOX gene family codes for transcriptional regulators with a variety of functions in plant developmental and physiological processes. In this study, a genome-wide comparative analysis of KNOX genes in Poplar (*Populustrichocarpa*) and rice (*Oryza sativa* L. ssp. japonica) was carried out. With comprehensive computational analyses, which take into account the gene structures, phylogeny and conserved motifs, 15 and 13 KNOX genes in Poplar and rice were identified, respectively. These KNOX genes were further divided into 3 groups. The Poplar gene POPTR_0012s04040 and the rice genes LOC_Os03g47042 and LOC_Os03g47022 were classified to a new group of KNOX genes without ahomeobox domain together with KNATM, which were proposed to play potential role in plant development and pluripotency. The identification of KNATM homolog in monocotyledons (rice) provided a strong support for proposing an ancient shuffling of *HOMEOBOX* gene with *MEINOX* gene took place in the KNOX phylogeny. Using subcellular location information, GO (gene ontology) and expression profile analysis, KNOX genes in rice and poplar were proposed to function similarly to the members in Arabidopsis. Our observations may lay the foundation for future functional analysis of KNOX genes in rice and poplar to unravel their biological roles in cellular pluripotency.

Keywords

KNOX Family, Evolutionary Expansion, *Oryza sativa*, *Populustrichocarpa*

1. Introduction

Homeobox genes are key regulators in plant and animal development. In plants, homeobox genes are divided into several groups according to their sequences, one of which is the KNOX (*knotted*-like homeob*ox*) family [1] [2]. Plant KNOX genes encode homeodomain-containing transcription factors that are required for meristem maintenance and organ initiation. There are 4 conserved motifs in KNOX proteins, KNOX1, KNOX2, ELK and Homeobox_KN [3] [4] [5]. KNOX1 and KNOX2, the two upstream domains collectively called MEINOX, are separated by a poorly conserved linker sequence [6].

In *Arabidopsis*, members of the KNOX genes are divided into three classes depending on their sequences and expression patterns, class I, class II and class M [7]. Class I subfamily consists of 4 members, SHOOT MERISTEMLESS (STM), KNAT1, KNAT2 and KNAT6 which are characteristically expressed in the shoot apical meristem (SAM). They play important roles in meristem, control of leaf shape and hormone homeostasis and act as either transcriptional activators or repressors [8] [9]. STM functions in shoot apical meristem, maintains and also regulates inflorescence architecture [10], KNAT1 also shows cell type specific expression patterns in the *Arabidopsis* root [11] and acts in a redundant fashion with STM [12] [13] [14]. KNAT2 is expressed during embryogenesis and marks the base of the SAM [15] [16], active in root tissue [17]. KNAT6 is expressed in the embryonic SAM, the SAM boundaries [15] and the phloem tissue of roots [18].

By contrast, the functions of the four *Arabidopsis* KNAT Class II genes *KNAT*3, *KNAT*4, *KNAT*5, and *KNAT*7, are barely known, but they are more widespread [8] [9]. KNAT3, KNAT4 and KNAT5 show cell type specific expression patterns in the *Arabidopsis* root [11]. The expression of KNAT3 is the highest in young siliques, inflorescences and roots, and the strongest expression of KNAT4 is in leaves and young siliques. Both of them show reduced expression in etiolated seedlings [19]. KNAT5 is expressed in the young primordium and the newly developing elongation zone of roots and its' expression in the epidermis marks the boundary between cell division and elongation [11]. KNAT7 is highly expressed in the central part of the *Arabidopsis* root [11] [20] and concerts with secondary wall formation in *Arabidopsis* and Poplar [21].

Class M represented by KNATM, a novel class of KNOX family in *Arabidopsis*, is characterized as missing the homeodomain by [7], and has a role in leaf proximal-distal patterning and is expressed in proximal-lateral domains of organ primordia at the boundary of mature organs [7].

Compared to the largely investigated functions of *Arabidopsis* KNOXs, three Poplar KNOX genes (*ARBORKNOX*1, *ARBORKNOX*2 and *PoptrKNAT*7) and nine rice KNOX genes (*Oskn*1/*HOS*59, *Oskn*2/*OSH*1, *Oskn*3/*OSH*15, *OSH*3, *OSH*6, *OSH*43, *HOS*58, *OSH*71 and *HOS*66) have been characterized up to date. *ARBORKNOX*1 (*ARK*1) and *ARBORKNOX*2 (*ARK*2) are both belong to Poplar Class I KNOX homeobox genes, and ortholog to *Arabidopsis STM* gene and

KNAT1 respectively. *ARK1* is expressed in the SAM and the vascular cambium, and is down-regulated in the terminally differentiated cells of leaves and secondary vascular tissues derived from these meristems [22]. ARK2 is expressed widely in the cambial zone and in terminally differentiating cell types, influences terminal cell differentiation and cell wall properties during secondary growth [23]. Three rice KNOX genes, *Oskn*1, *Oskn*2 and *Oskn*3, are expressed in the SAM and are involved in regulating SAM formation [24], *OSH*3 is expressed in the inflorescence meristem and involved in morphogenesis [25] [26]. *HOS*66 has a broadly expression from root to leaf, flower and callus [27] while its' function is still unraveled.

Though a survey and classification of KNOX genes in maize had been carried out 20 years ago [1] and expression analyses [25] [28] and function analysis of a few rice KNOXs [24] [29] have been carried out, no detailed systematic analysis including subcellular location, gene structure and expression profiling has been conducted. In order to reveal the evolution and function of KNOX genes in monocot and dicot plants further, a genome-wide identification of KNOX genes in Poplar and rice were taken as reference for dicot and monocot, respectively. The sequence phylogeny, genome organization, gene structure, conserved motifs and expression profiling of those homologous genes were also analyzed to verify the evolution origins of Poplar KNOX genes as well as to confirm their tissue-specific expression patterns. This might provide some insights for future engineering modifications of plant pluripotency and regeneration characteristics in rice and Poplar.

2. Material and Methods

2.1. Database Search and Sequence Retrieval

Multiple database searches were performed to collect all members of the rice (*Oryza sativa*) and Poplar (*Populustirchocarpa*) KNOX family. The amino acid sequence of the 9 KNOX genes from *Arabidopsis* (*Arabidopsis thaliana*) were used as query sequences, BLAST programs (TBLASTN and BLASTP) on the Phytozome (http://www.phytozome.net/), TAIR (http://www.arabidopsis.org/), Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/), NCBI (http://www.ncbi.nlm.nih.gov/) and PlantDB (http://www.plantgdb.org/) genome database were performed and the sequences hitting with 1e−6 or less were treated as conserved genes. To increase the extent of the database search results, position-specific iterated BLAST [30] search against the rice and Poplar database on the NCBI web site were also performed. All the sequences with conserved domains of KNOX protein except for those annotated as miRNA, retrotransposons or transposable elements were considered as KNOX candidates.

2.2. Construction of the Phylogenetic Tree and Gene Structure Analysis

To generate the phylogenetic tree of each gene family, a multiple alignment

analysis was performed with ClustalW and the rooted phylogenetic tree with branch length was constructed by UPGMA (Unweighted Pair Group Method with Arithmetic mean) method with amino acid sequences. In UPGMA the unweighted term indicates that all distances contribute equally to each average that is computed and does not refer to the math by which it is achieved and the proportional averaging in UPGMA produces an unweighted result. The UPGMA algorithm constructs a rooted tree (dendrogram) that reflects the structure present in a pairwise similarity matrix (or a dissimilarity matrix). At each step, the nearest two clusters are combined into a higher-level cluster. The distance between any two clusters A and B, each of size |A| and |B| is taken to be the average of all distances d(x, y) between pairs of objects x in A and y in B, that is, the mean distance between elements of each cluster. The UPGMA algorithm produces rooted dendrograms and requires a constant-rate assumption that is, it assumes an ultrametric tree in which the distances from the root to every branch tip are equal.

Gene structure display server (GSDS) program (http://gsds.cbi.pku.edu.cn/) [31] was used to illustrate exon/intron organization for individual KNOX genes by aligning the coding sequences (CDS) with their corresponding genomic sequences from Phytozome (http://www.phytozome.net). The visualization of gene features such as composition and position of exons and introns for genes offers visual presentation to integrate annotation, and to produce high-quality figures.

2.3. Identification and Annotation of Conserved Motifs

The program MEME (v4.9.0) (http://meme.nbcr.net/meme) [32] and ClustalW multiple alignment analyses (http://www.genome.jp/tools/clustalw) were used for the elucidation of conserved motifs in all deduced rice and Poplar KNOX protein sequences. The following parameters were used: number of repetitions any, maximum number of motifs 20, and the optimum motif widths were constrained to between 6 and 60 residues. The HMM (Hidden Markov Models) logos were used for visualization of domain conservation. Structural motif annotation was performed using the SMART (http://smart.embl-heidelberg.de) and Pfam (http://pfam.sanger.ac.uk) databases.

To identify the number of domains in KNOX protein, domain search were executed by Conserved Domains Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [33] and pfam database (http://pfam.sanger.ac.uk/) [3] with both local and global search strategy and expectation cut off (E value) 1.0 was set as the threshold for significance. Only significant domain found in rice and Poplar KNOX protein sequences were considered as a valid domain.

2.4. Protein Structure Prediction

The 3D (three-dimensional structure) protein structure prediction of KNOXs in

Arabidopsis, rice and Poplar were performed by PHYRE2 (Protein Homology/ AnalogY Recognition Engine) Protein Fold Recognition Server (www.sbg.bio.ic.ac.uk/phyre2/) [34]. PHYRE2 is web-based services for protein structure prediction using the principles and techniques of homology modeling. It is able to regularly generate reliable protein models when other widely used methods such as PSI-BLAST cannot.

2.5. GO Analysis and Subcellular Localization

All targets identified in this study were subjected to agriGO toolkit analysis (http://bioinfo.cau.edu.cn/agriGO) to investigate gene ontology [35].

The subcellular localization of KNOX proteins were predicted by SVM based server ESL pred (http://www.imtech.res.in/raghava/eslpred/submit.html), and ProtComp 9.0 server (http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=help&subgroup=proloc). Further, species-specific localization prediction system was utilized for *Arabidopsis* (AtSubP, http://bioinfo3.noble.org/AtSubP) [36].

2.6. Expression Analysis for KNOX Genes in Poplar and Rice

For Poplar, the expression profile for each *KNOX* gene was obtained by evaluating its EST representation among 19 cDNA libraries derived from different tissues and/or developmental stages available at PopGenIE (http://www.popgenie.org/) [37]. The absolute mode was used.

For rice, expression support for each gene model was explored through gene expression evidence search page (http://rice.plantbiology.msu.edu/expression.shtml) available at MSU. Expression data were derived from NCBI Sequence Read Archive (SRA). Sequence reads were mapped to the version 7 pseudomolecules with Tophat. Expression abundances for RNAseq libraries were calculated with Cufflinks [38] [39]. The Sequence Read Archive (SRA) stores raw sequencing data from the next generation of sequencing platforms including Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System, HelicosHeliscope®, Complete Genomics®, and Pacific Biosciences SMRT®.

3. Results

3.1. Identification and Phylogenetic Analysis of the KNOX Family Genes in Poplar and Rice

By BLAST(BLASTp and BLASTn) search against multiple databases with query sequences of nine previously published KNOX genes from *Arabidopsis*, the candidate KNOX genes in rice and Poplar were further analyzed to confirm the presence of KNOX1, KNOX2, ELK and Homeobox_KN domain through Pfam program and Conserved Domain Search Service (CD Search) in NCBI. Excluding the redundant sequences, 13 rice (*Oryza sativa*. L) KNOX genes and 15 Poplar (*populoustirchocarpa*) KNOX genes were identified (**Table 1**).

According to the conserved amino acid sequences, all these KNOXs genes in

Table 1. KNOX gene information[a].

Gene	Accession number	Genomic sequence (bp)	Transcript sequence (bp)[b]	CDS sequence (bp)	Protein sequence (AA)	Species
KNATM	AT1G14760	526	429	429	142	*Arabidopsis thaliana*
KNAT6	AT1G23380	5310	1306	990	329	*Arabidopsis thaliana*
STM	AT1G62360	3141	1363	1149	382	*Arabidopsis thaliana*
KNAT7	AT1G62990	3324	1319	876	291	*Arabidopsis thaliana*
KNAT2	AT1G70510	5946	1368	933	310	*Arabidopsis thaliana*
KNAT1/BP	AT4G08150	3267	1822	1197	398	*Arabidopsis thaliana*
KNAT5	AT4G32040	2292	1561	1152	383	*Arabidopsis thaliana*
KNAT4	AT5G11060	3188	1810	1182	393	*Arabidopsis thaliana*
KNAT3	AT5G25220	2511	1899	1296	431	*Arabidopsis thaliana*
OSH6	LOC_Os01g19694	7505	1619	906	301	*Oryza sativa. L*
HOS58	LOC_Os02g08544	3474	1576	942	313	*Oryza sativa. L*
HOS66	LOC_Os03g03164	5107	1468	945	314	*Oryza sativa. L*
	LOC_Os03g47022	925	501	501	166	*Oryza sativa. L*
	LOC_Os03g47042	1274	585	585	194	*Oryza sativa. L*
	LOC_Os03g51690	10585	2522	1086	361	*Oryza sativa. L*
OSH3	LOC_Os03g51710	6460	2220	1134	377	*Oryza sativa. L*
OSH43	LOC_Os03g56110	5946	1844	1026	341	*Oryza sativa. L*
	LOC_Os03g56140	5078	1350	1158	385	*Oryza sativa. L*
OSH71	LOC_Os05g03884	6492	1951	936	311	*Oryza sativa. L*
HOS59/Oskn1	LOC_Os06g43860	3348	1584	972	323	*Oryza sativa. L*
OSH15/Oskn3	LOC_Os07g03770	6124	1628	1068	355	*Oryza sativa. L*
OSH1/Oskn2	LOC_Os08g19650	6566	1685	837	278	*Oryza sativa. L*
PoptrKNAT7	POPTR_0001s08550	5065	1806	906	301	*Populustrichocarpa*
ARBORKNOX2	POPTR_0002s11400	6341	2225	1107	368	*Populustrichocarpa*
	POPTR_0004s00650	4369	1475	1110	369	*Populustrichocarpa*
	POPTR_0005s01720	5189	1738	954	317	*Populustrichocarpa*
	POPTR_0006s20440	5506	1657	1017	338	*Populustrichocarpa*
	POPTR_0006s27560	5145	2117	1299	432	*Populustrichocarpa*
	POPTR_0008s19300	8346	1474	1026	341	*Populustrichocarpa*
	POPTR_0010s05340	10307	1486	930	309	*Populustrichocarpa*
ARBORKNOX1	POPRT_0011s01600	4373	1319	1122	373	*Populustrichocarpa*
	POPTR_0012s04040	4910	546	546	181	*Populustrichocarpa*
	POPTR_0012s08910	8044	1474	1023	340	*Populustrichocarpa*
	POPTR_0013s01000	4664	1458	963	320	*Populustrichocarpa*
	POPTR_0015s09030	8001	1690	1044	347	*Populustrichocarpa*
	POPTR_0018s02210	5133	2209	1281	426	*Populustrichocarpa*
	POPTR_0018s12200	4420	1652	1005	334	*Populustrichocarpa*

[a]:The sequence information was from Phytozome database (http://www.phytozome.net/); The longest isoform was chosen if there were more than one alternative splicing isoform available.

Arabidopsis, rice and Poplar divided into three groups, class I, class II and class III, which agreed with representative genes of *Arabidopsis* (Figure 1). Among which, 22 KNOX genes belonged to class I, including 4 *Arabidopsis* KNOX genes, 9 *Popolus* KNOX genes and nine rice KNOX genes, 13 genes belonged to class II and 2 genes belonged to class III.

Class I was further divided into 3 subclasses, designated as IA, IB and IC. Class IC comprises only 4 rice genes, LOC_Os03g56140, LOC_Os03g56110, LOC_Os03g47022 and LOC_Os03g47042, showed ambiguous clustering from other phylogenetic classes (Figure 1).

Class II included IIA and IIB two subclasses. KNAT7, together with rice gene LOC_Os03go3164 and Poplar member POPTR_0001s08550 represented class IIA subclass. Besides *Arabidopsis* KNAT3, KNAT4 and KNAT5, 3 rice members, LOC_Os08g19650, LOC_Os02g08544 and LOC_Os06g43860 and 4 Poplar members, POPTR_0006s20440, POPTR_0018s12200, POPTR_0006s27560, and POPTR_0018s02210 were classified into class IIB clade (Figure 1).
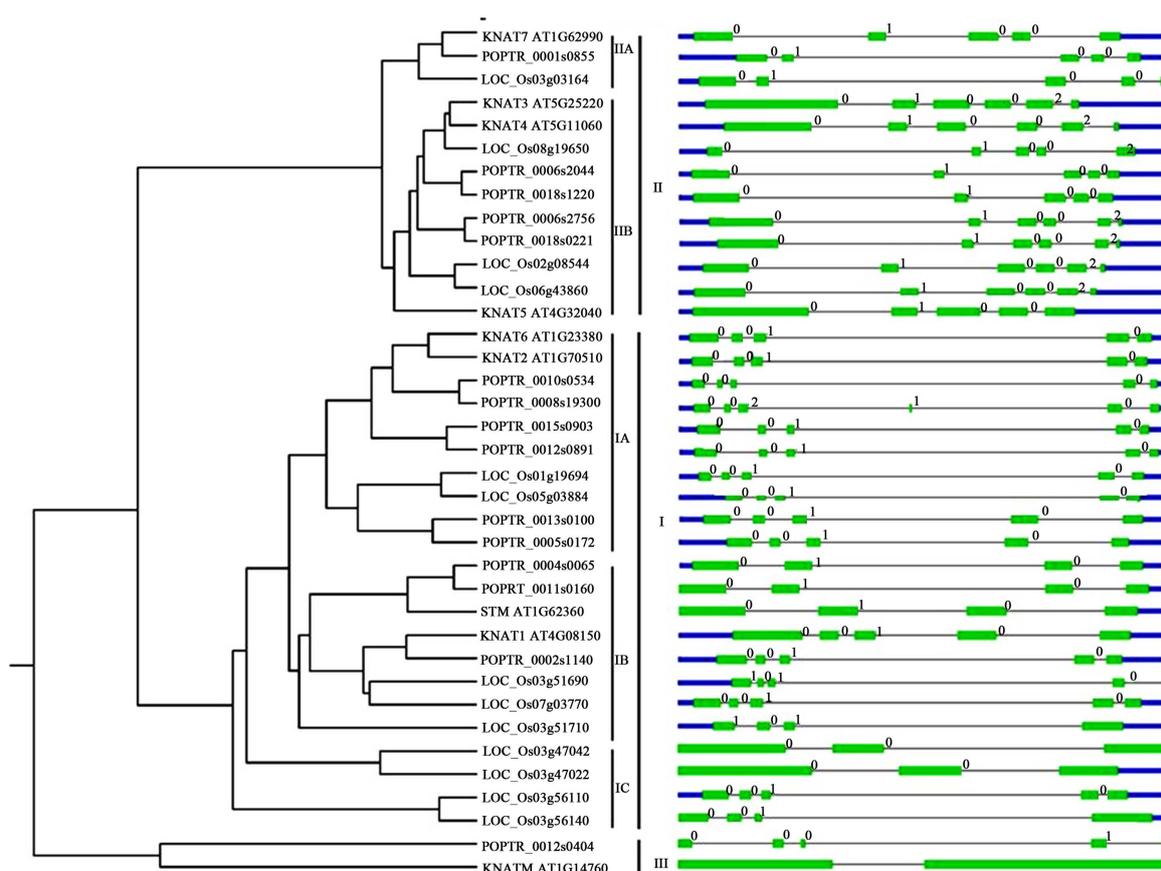
Figure 1. Phylogenetic relationship and gene structure of the KNOX members in *Arabidopsis*, rice and *Populus*. The phylogenetic tree was constructed with the amino acid sequences of KNOX proteins from Rice (*Oryza sativa*) (LOC_Os), *Arabidopsis* (*Arabidopsis thaliana*) (AT) and *Populus* (*Populustrichocarpa*) (POPTR) using UPGMA (rooted phylogenetic tree with branch length) method by Clustalw. The gene structure was performed by GSDS program with the sequences from Phytozome. exon/intron structures are represented by the green bars which showing exons and grey lines which showing introns. The blue bar showing 5' and 3' UTR, number 0, 1, 2 corresponds to the intron phase. The sizes of exons and introns are proportional to their sequence lengths.

Poplar KNOX gene POPTR_0012s04040 fell into class III together with *Arabidopsis* KNOX member KNATM, showed big divergences from the other clades and indicated that class III was new branch in the phylogeny (**Figure 1**).

3.2. Comparison of Gene Structure in *Arabidopsis,* Poplar and Rice

Structural diversity of KNOXs in rice and Poplar analyzed based on the alignment of cDNA and corresponding genomic sequences by GSDS revealed that KNOX members within the same subfamilies shared very similar gene structure in terms of either intron numbers or exon lengths except for class IC.KNOXs in class IA shared a structure with 5 exons and the third and fourth exons were separated by a long phase 1 intron. According to the sequences from Phytozome, class III member KNATM had only one intron, whether it had 5' and 3' UTR still need to be identified (**Figure 1**). Nonetheless, the gene structures in 4 rice members of subclass IC appeared to be more variable. LOC_Os03g56110 had a same structure as subclass IA, LOC_Os03g56140 had 4 exons and a long phase 1 intron interval between the third and fourth exon. Both of LOC_Os03g47042 and LOC_Os03g47022 had 3 exons separated by two phase 0 introns, the 5' UTR and 3'UTR of gene LOC_Os03g47042 were not available and the gene LOC_Os03g47022 had 3'URT only, that showed similar features with KNATM in gene structure.

In addition, the intron phases were highly conserved in KNOX genes within each class. Class I and Class III always was in phase 0 and 1, while 50% genes in class II was in phases 2 and 0 (**Figure 1**), resulting in a significant excess of non-symmetrical exons. That suggested splicing phases were also highly conserved during the evolution and there was a strong correlation between the phylogeny and exon/intron structure of the KNOX gene family in rice and Poplar.

Although the intron phases with respect to codons were remarkably well-conserved within the same subfamilies, there were striking distinctions in the arrangement of introns and intron phases among subfamilies of KNOX IB (**Figure 1**). The conservation of intron phases within KNOX subfamilies and the striking dissimilarity between subfamilies may reciprocally lend support to the results from phylogenetic analysis and genome duplication, and provide further supports to plant KNOX subfamilies definition.

3.3. Multiple Alignment and Conserved Domain Analysis of KNOXs in Each Class

The domain conservation of 13 rice and 15 Poplar KNOX proteins were analyzed by performing a multiple sequence alignment with *Arabidopsis* KNOX amino acid sequences by ClustalW (**Figure S1**, **Figure S2**, **Figure S3**) and MEME program. No KNOX2 domain was found in rice gene LOC_Os03g56140. Both ELK domain and Homeobox_KN domain were not found in rice sequences LOC_Os03g47042 and LOC_Os03g47022 (**Figure 2**). These two genes showed similar characters as class III member, *Arabidopsis* KNATM and POPTR_0012s04040, which characterized as no ELK domain and Homeobox_KN domain (**Figure 2**).
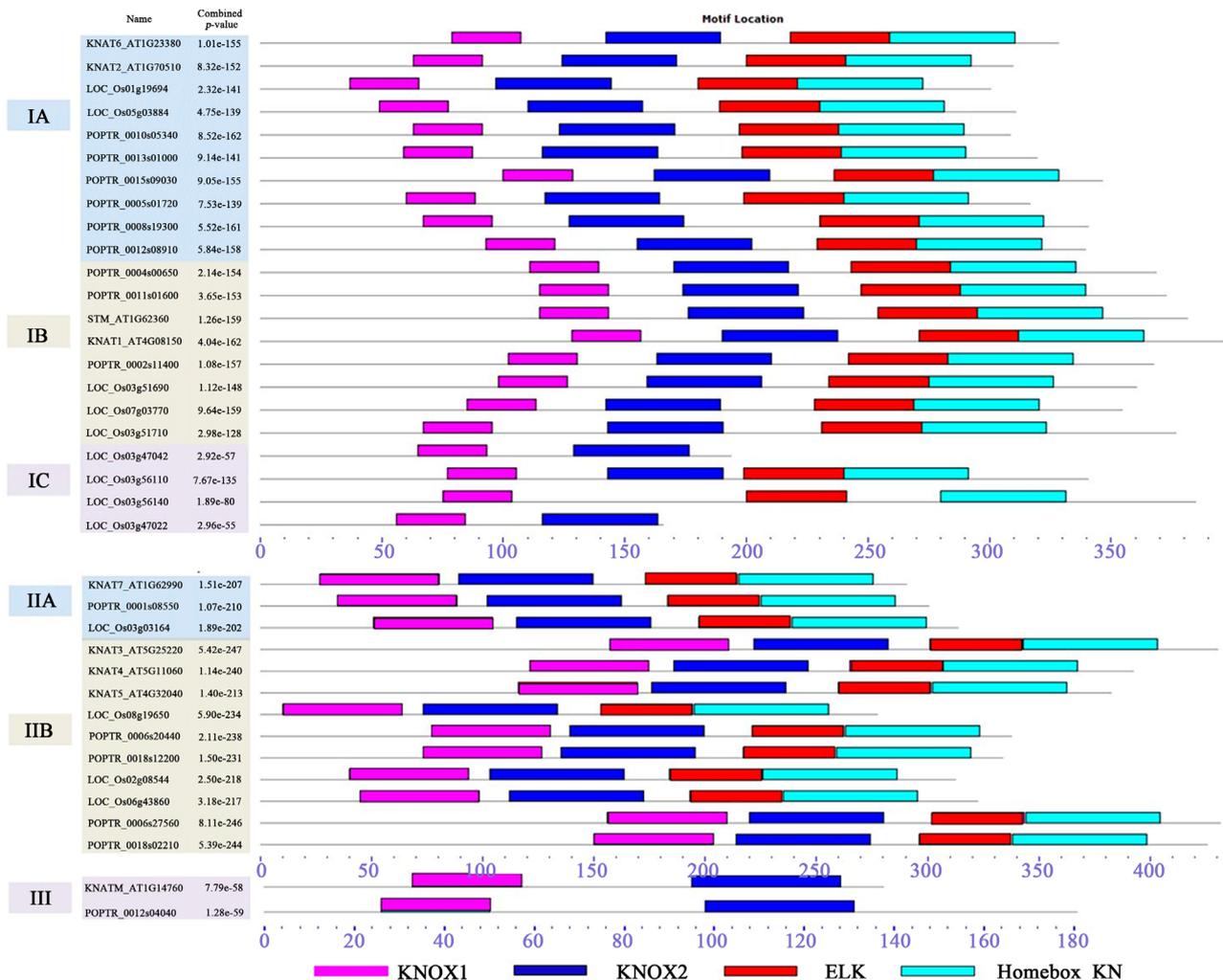
Figure 2. Conserved domain analysis of KNOXs with MEME program. Based on domain conservation, all these KNOXs genes in Arabidopsis, rice and Poplar were divided into three groups, class I, class II and class III. Class I was further divided into 3 sub-classes, designated as IA, IB and IC. Class II was further divided into subclasses IIA and IIB. Combined P-value and motif locations were showed. Non-overlapping sites with a p-value better than 0.0001. The height of the motif "block" is proportional to –log (p-value), truncated at the height for a motif with a p-value of $1e^{-10}$.

For other class I genes and class II genes, KNOX1, KNOX2, ELK and Homeo-box_KN domains were highly conserved (Figure 2). The KNOX1 domain was completely conserved in all of the 22 class I sequences and the E-Value was as high as 4.3e-309 (**Figure S1**).

The domain variation of three genes in class IC showed consistence with their divergence in phylogenic analysis. And the domain similarities among LOC_Os03g47042, LOC_Os03g47022, KNATM and POPTR_0012s04040 suggested that the gene cluster analysis sometimes cannot classify the high divergence of the motif, especially combined with its short length and the occurrence of many ancient paralogs when large number of sequences being examined, which can be complementary by domain analysis. Same results were also obtained through CD (conserved domain) search in NCBI and Pfam (**Table S1**).

3.4. Protein 3D Structure Predication of KNOX Family Genes in Rice and Poplar

The 3D protein structures of KNOXs in rice and Poplar predicated by PHYRE2 showed variances within different subclasses. All of the members in class IA, class IB and class II members had highly conserved 3D structures (**Figure 3(a)**) featured 3 helices, among which helices I and II were typically connected by a loop structure, helices II and helices III formed a helix-turn-helix motif. The structure of KNAT4 was a little different which included a set of 3 helices structure at N terminal and a set of 5 helices structure at C terminal (**Figure 3(b)**). A bipartite arrangement of these two set helix structure was separated by a linker region. This kind of structure configuration hinted the functional diversity of KNAT4 in *Arabidopsis* development.
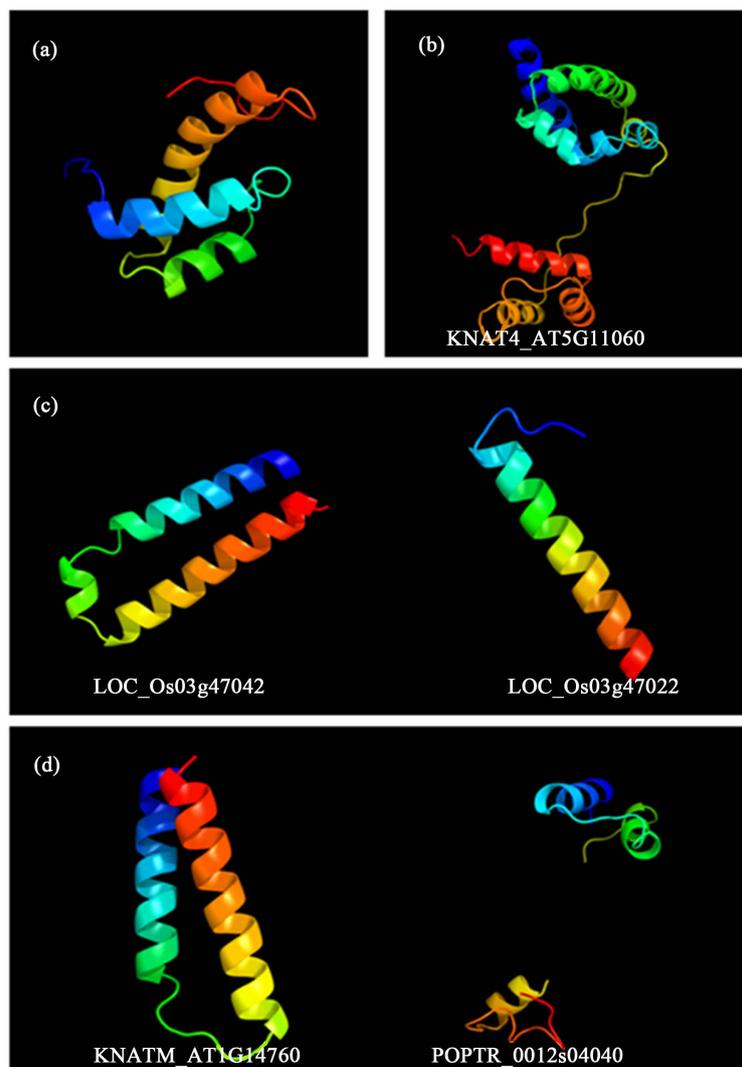
Figure 3. The 3D protein structure of KNOX genes. (a): Typical 3 helices structure of KNOX genes in this study; (b): 3D structure of KNAT4; (c): 3D structure of two class IC KNOX genes LOC_Os03g47042 and LOC_Os03g47022; (d): 3D structure of class M. Image coloured by rainbow N → C terminus.

However, the 3D structures of 4 rice gene members in class IC showed high variances. LOC_Os03g56110 and LOC_Os03g56140 still had typical 3 helices structure like other KNOX genes, while rice gene LOC_Os03g47042 had very short helices II, no typical loop structure between helices I and helices II and no helix-turn-helix motif formed by helices II and helices III, and LOC_Os03g47022 showed merged long helices, no loop, turn and other two helices (Figure 3(c)). The typical 3D protein structures against the 3-helix-structureagreed with the protein structure prediction of class III KNOXs (Figure 3(d)). *Arabidopsis* gene KNATM (AT1G14760) had a structure with two longer helices connected by a short linker, *populous* gene PORT_0012s04040 had a structure with unlinked helices, the first helices was disconnected with the rest two. The conservation and variances of 3D protein structure in class I was highly in accordance with phylogenic analysis and conserved domains analysis, which further supported they represented highly diverged lineage-specific KNOX sequences. These a typical 3D protein structures of LOC_Os03g47022, LOC_Os03g47042 and PORT_0012s04040 implied their special biological function and phylogenic divergence.

3.5. Differential Expression Profile and Subcellular Location of KNOX Family Genes in Poplar and Rice

To understand the temporal and spatial expression patterns of rice and Poplar KNOX genes, we compared their expression patterns during development with the expression data derived from RNA-seq available at MUS and from PopGenIE respectively.

The diversity expression patterns of rice KNOX genes in 11 different tissues showed that except unavailable expression data for LOC_Os03g47042, most of the class I genes mainly expressed in pre-emergence inflorescence, post-emergence inflorescence, and pistil, and 25 DAP (days after pollination) embryo with different expression dynamics (Table 2), gene LOC_Os05g03884had highest expression in pistil and gene LOC_Os01g19694in 25 DAP embryo. LOC_Os07g03770 also had a significant expression in anther besides the highest expression in pre-emergence inflorescence with expression value as high as 52.75 (Table 2). Comparing to other class I genes, the gene LOC_Os03g51710 had expression in all of 11 tissues and its highest expression was in pre-emergence inflorescence, the value reached 5.86 (Table 2).

All of the 4 class II genes almost expressed in all of the 11 tissues which were analyzed and had the same expression pattern with strong expression in shoot, 4-leaf stage seedling and 20 days leaf, and the gene LOC_02g08544 showed a ubiquity expression (Table 2).

For Poplar, the 5 class I KNOX genes (POPTR_0008s19300, POPTR_0013s01000, POPTR_0005s01720, POPTR_0004s00650, and POPTR_0002s11400) had strong expression in root, mature leaf, internodes, young leaf and nodes, while no expression of gene POPTR_0010s05340, POPTR_0015s09030, POPTR_0012s08910, POPRT_0011s01600 and Class III gene POPTR_0002s11400 were available in these five tissues. The gene POPTR_0008s19300 had a stronger expression in

Table 2. Rice expression[a].

	Locus_id	T1[b]	T2[b]	T3[b]	T4[b]	T5[b]	T6[b]	T7[b]	T8[b]	T9[b]	T10[b]	T11[b]	Subcellular location
	LOC_Os01g19694	0.73	0.00	0.00	0.13	0.00	3.37	9.94	0.00	3.40	10.04	0.00	Nucleus
	LOC_Os05g03884	6.55	0.00	0.00	0.15	0.00	12.12	8.51	0.17	15.92	11.21	0.00	Nucleus
	LOC_Os07g03770	1.43	0.00	0.00	0.24	0.00	28.77	52.75	1.69	13.02	4.92	0.00	Unknown
	LOC_Os03g51690	1.86	0.00	0.00	0.14	0.00	65.14	80.58	0.21	48.61	9.46	0.00	Unknown
Class I	LOC_Os03g51710	3.13	1.75	1.66	3.01	1.66	5.05	5.86	1.79	5.51	0.16	0.88	Unknown
	LOC_Os03g56140	0.00	0.00	0.00	0.00	0.00	2.15	1.77	0.39	1.19	0.00	0.00	Nucleus
	LOC_Os03g56110	0.76	0.43	0.00	0.05	0.00	12.97	2.75	3.15	1.86	0.00	0.00	Nucleus
	LOC_Os03g47022	0.00	0.00	3.85	4.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Nucleus
	LOC_Os03g47042	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	Nucleus
	LOC_Os06g43860	7.06	0.68	166.19	194.09	197.13	59.28	10.34	61.46	13.25	1.10	0.97	Nucleus
Class II	LOC_Os08g19650	2.86	0.00	53.06	46.38	35.25	8.87	3.67	3.79	21.10	1.14	0.00	Nucleus
	LOC_Os03g03164	0.95	0.45	56.38	14.00	0.17	12.16	9.38	4.22	4.09	0.00	0.76	Nucleus
	LOC_Os02g08544	19.65	4.96	60.74	60.42	63.67	48.44	42.98	22.48	46.13	16.93	4.87	Nucleus

[a]: expression support for each gene model was explored through gene expression evidence search page (http://rice.plantbiology.msu.edu/expression.shtml) available at MSU. Expression data were derived from NCBI Sequence Read Archive (SRA). [b]: T1: Seed-5 DAP; T2: Seed-10 DAP; T3: Shoots; T4: Seedling four-leaf stage; T5: Leave 20 days; T6: Post-emergence inflorescence; T7: Pre-emergence inflorescence; T8: Anther; T9: Pistil; T10: Embryo-25 DAP; T11: Endosperm-25 DAP.

nodes than in other four tissues. The gene POPTR_0013s01000 had stronger expression in both of internodes and nodes with the expression value of 9.37 and 9.47 respectively. The expression of gene POPTR_0005s01720, gene POPTR_0004s00650 and gene POPTR_0002s11400 in root, internodes and nodes were higher than in other tissues (Table 3). The class II KNOX genes showed evenly expression in root, mature leaf, internodes, young leaf and nodes with highest expression in mature leaf (Table 3).

Protein subcellular localization is crucial for protein function prediction. The subcellular location of KNOX members in rice (Table 2) and Poplar (Table 3) showed that except for three rice class IB members were still unknown, all others were located in nucleus, which implied their transcription factor functions.

3.6. GO (Gene Ontology) Analysis of KNOX Target Genes in Rice and Poplar

To identify biological processes these KNOXs might participate in and whether they were differential in rice and Poplar, 15 Poplar KNOX genes and 13 rice KNOX genes were subjected to AgriGO toolkit analysis to investigate gene ontology. 13 rice genes were all involved in cellular process, regulation of biological process, biological regulation, metabolic process, transcription regulator activity and binding activity (Figure 4(a)). All 15 Poplar KNOXs genes were involved in DNA binding activity and more than 70% of the Poplar genes were involved in biological regulation and transcription regulator activity (Figure 4(b)). The
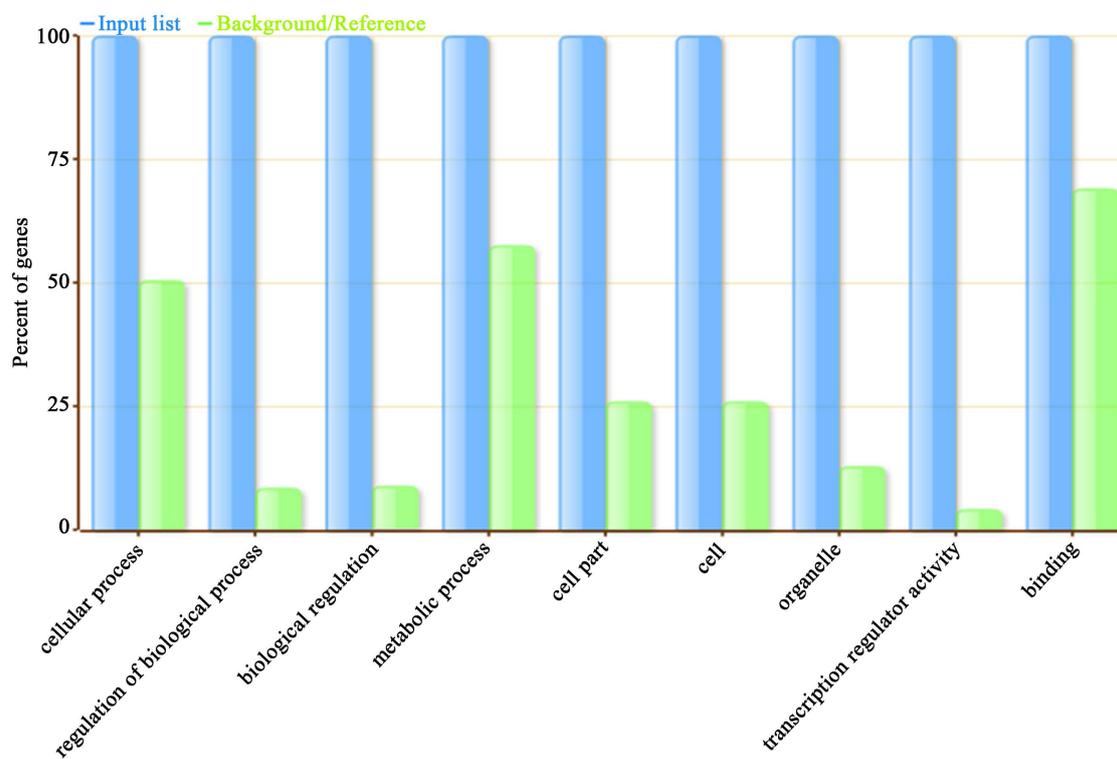
(a)
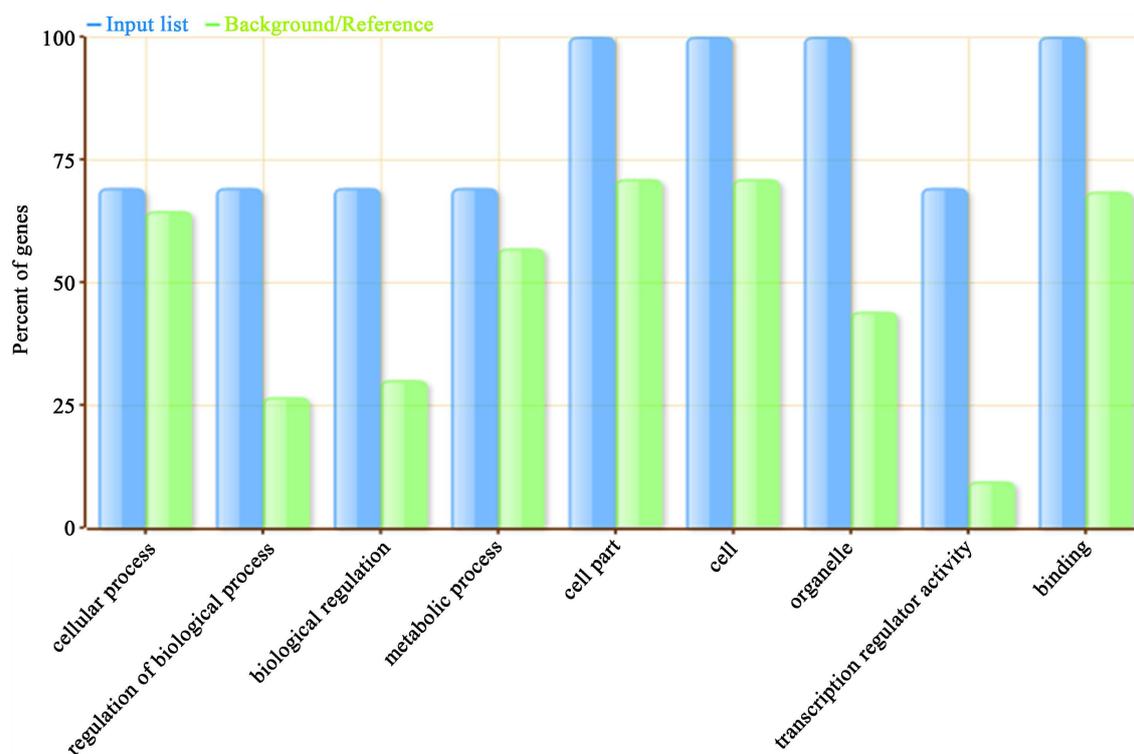
(b)

Figure 4. GO flash chart of biological process and molecular function by GO analysis of targets of KONXs in rice (a) and *Populus* (b). Blue bars indicate the enrichment of *Populus* KNOXs targets in GO terms. Green bars indicate the percentage of total annotated *Populus* genes mapping to GO terms.

Table 3. Expression and subcellular location of Populus KNOX genes[a].

	Gene	Roots	Mature-leaves	Internodes	Young-leaves	Nodes	Subcellular location
	POPTR_0010s05340	0	0	0	0	0	Nucleus
	POPTR_0008s19300	6.36	5.72	5.83	5.81	7.23	Nucleus
	POPTR_0015s09030	0	0	0	0	0	Nucleus
	POPTR_0012s08910	0	0	0	0	0	Nucleus
Class I	POPTR_0013s01000	8.38	6.63	9.37	6.83	9.47	Nucleus
	POPTR_0005s01720	8.34	6.89	8.42	7.33	8.37	Nucleus
	POPTR_0004s00650	9.53	7.24	9.85	6.87	9.77	Nucleus
	POPRT_0011s01600	0	0	0	0	0	Nucleus
	POPTR_0002s11400	12.07	7.76	12.87	8.41	12.88	Nucleus
	POPTR_0001s08550	9.83	10.08	9.78	8.94	9.07	Nucleus
	POPTR_0006s20440	8.63	9.68	7.91	9.53	8.6	Nucleus
Class II	POPTR_0018s02210	8.52	8.78	8.3	8.43	8.18	Nucleus
	POPTR_0006s27560	10.04	11.04	9.91	10.77	9.74	Nucleus
	POPTR_0018s12200	9.77	10.27	8.59	10.07	9.63	Nucleus
Class III	POPTR_0012s04040	0	0	0	0	0	Nucleus

[a]: the expression profile were obtained by evaluating its EST representation among 19 cDNA libraries derived from different tissues and/or developmental stages available at PopGenIE (http://www.popgenie.org/).

enrichment of rice and Poplar KNOXs involved in DNA binding was consistent with the subcellular predication that most of them were located in nucleus.
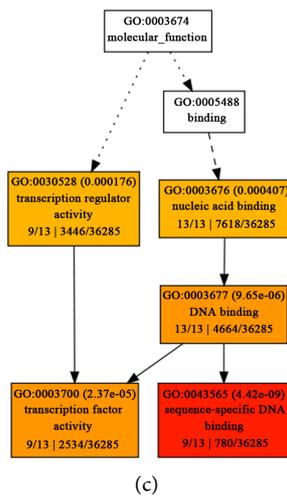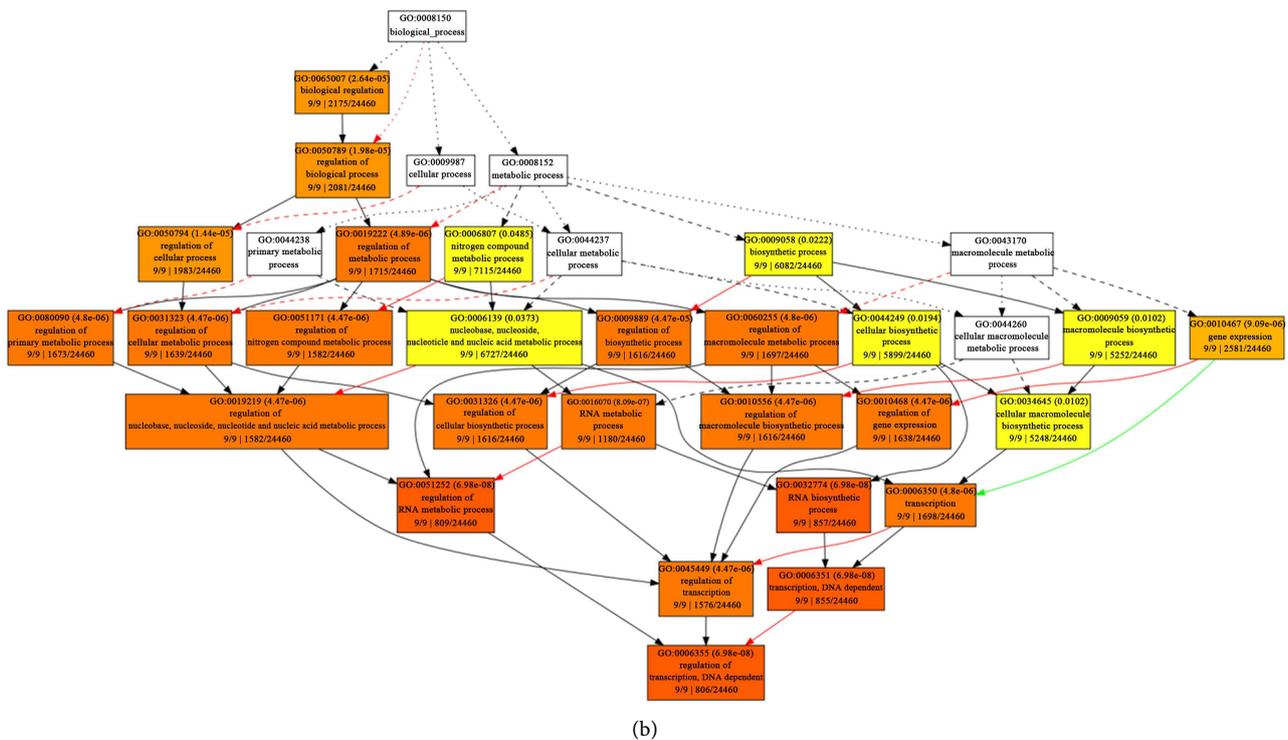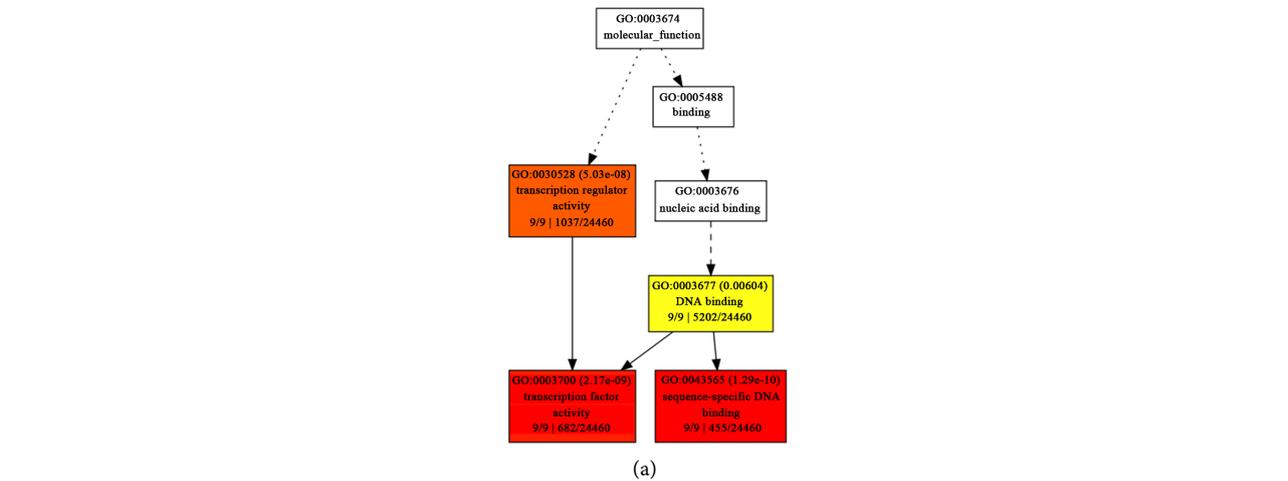
According to GO analysis, the functional group for transcription regulator activity, transcription factor activity and sequence specific DNA binding was highly enriched by KNOX target genes in rice and Poplar nucleus (Figure 5), but the enrichment of functional group for DNA binding and transcription factor activity was not as high as that of sequence-specific binding in Poplar. That showed all of rice KNOXs and most of Poplar KNOXs had roles in DNA binding and transcription regulator activity when considered their expression and subcellular analysis together.

Except for transcription factor activity and DNA binding, KNOXs target genes in rice were highly functionally related to RNA metabolic process, regulated RNA metabolic process and RNA biosynthesis process (Figure 5(a) and Figure 5(b)). It confirmed that rice KNOXs were involved in plant development as transcriptional factors. Poplar KNOXs target genes were weaker in regulation of RNA metabolic process and RNA biosynthesis process (Figure 5(c) and Figure 5(d)).

4. Discussion

4.1. Classification and Evolution of KNOXs in Rice and Poplar

The KNOX genes in *Arabidopsis* are divided into class I, class II and class M

(a)

(b)

(c)
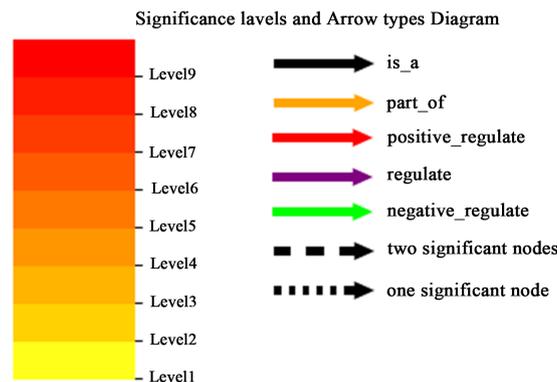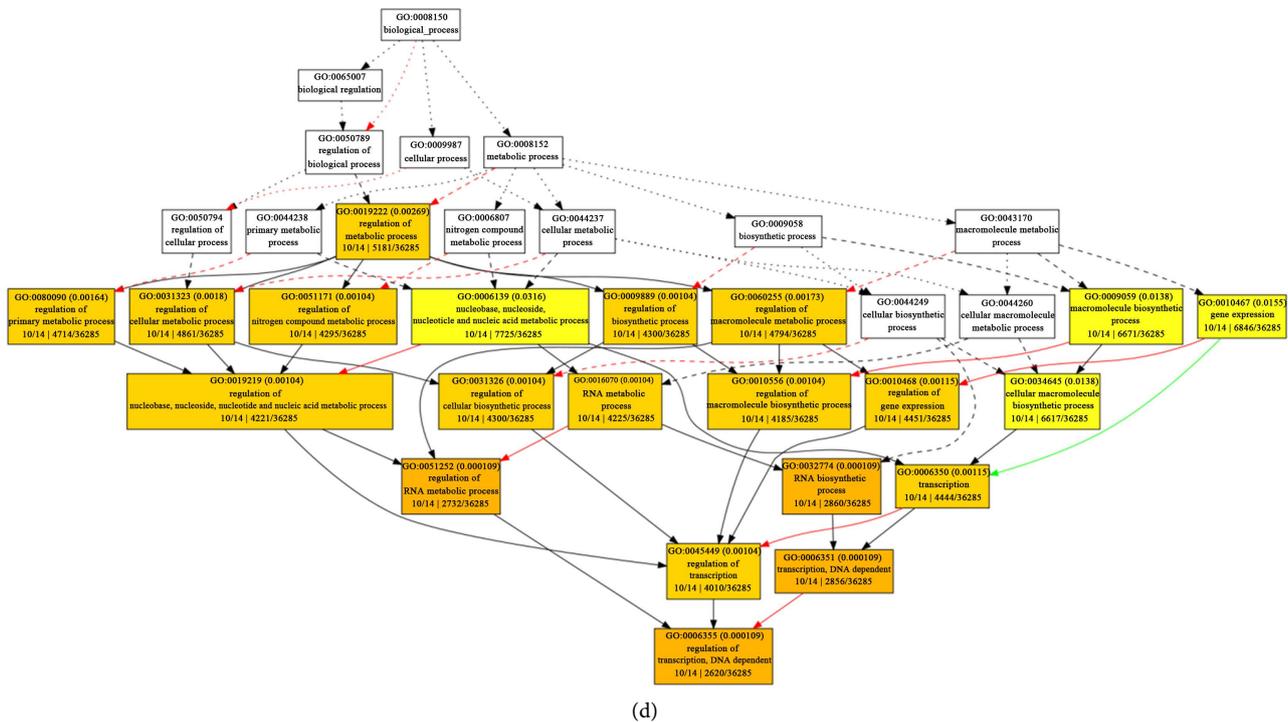
Figure 5. Gene Ontology (GO) analysis of the KNOX genes. GO analysis according to AgriGO (http://bioinfo.cau.edu.cn/agriGO/index.php), with respect to molecular function ((a), (c)) and biological process ((b), (d)) in rice ((a), (b)) and populous ((c), (d)). p-value was assigned to each GO group based on the overabundance of significant genes. The block color from yellow to red was divided in 9 levels to represent an increasing of enrichment strength roughly.

depending on their sequences and expression patterns [7]. According to sequences search and blast in this study, KNOXs homologous are scattered across rice and Poplar chromosome and can also be grouped into 3 classes based on sequences and gene structure (**Figure 1**). That indicated KNOXs were an ancient gene family and conserved highly across species. However, class IC consisted of only four rice genes, was divergent clustering between other phylogenetic classes. They most likely represented highly diverged lineage-specific KNOX sequences or the phylogenetic analysis could not resolve their evolutionary relationships.

Two rice genes, LOC_Os03g47016 and LOC_Os03g47036, were classified into KNOX gene family by Jain *et al.* (2008), while no KNOX1 and KNOX2 domains

were found by us, so they were excluded from KNOX family in this study. However, two new KNOX genes, LOC_Os03g47042, LOC_Os03g47022 were firstly identified in this study and grouped into class IC according to the amino acid sequences, but when considered the gene structure, conserved domain and 3D protein structure all together, they showed more similar to *Arabidopsis* class M members and might be the KNATM orthologs in rice. As deep nodes and determining interclade relationships commonly showed low statistical support and varied between different phylogenetic methods which typically observed in protein phylogenies [40], the ambiguous classifications indicated the gene cluster analysis sometimes cannot classify the high divergence of the motif, especially combined with its short length and the occurrence of many ancient paralogs when large number of sequences being examined, which can be complementary by domain analysis.

KNATM homolog was found only in dicots and placed in a new class in KNOX family [7]. The KNATM homolog in Poplar suggested that KNATM originated early in the evolution of dicotyledons. What's more, two possible KNATM homolog genes were found in rice (Figure 1), it was the first time to identify MEINOX protein in monocotyledon and the discovery argued against the hypothesis that the KNATM homolog in monocotyledons is a canonic KNOX protein with the homeodomain being redundant or KNATM originated in dicotyledons was lost in monocotyledons. All of these argue against the hypothesis of KNATM being a pseudogene.

4.2. Functions of KNOXs in Rice and Poplar

As key effectors involved in transcriptional regulation and hormonal signaling, the function diversity is performed by the KNOX members in different classes. Class IA, IB and Class II KNOXs in rice and Poplar all had 4 highly conserved domains and were located in nucleus, highly expressed in undifferentiated tissue (Table 2 and Table 3), functioned in transcriptional factor activity (Figure 5), that suggested them might participate in plant cell differentiation and plant morphogenesis, with similar functions as that in *Arabidopsis*.

In *Arabidopsis*, class I KNOX genes were mainly expressed in the meristematic tissues and regulate hormonal pathways to maintain meristematic cells in an undifferentiated state [5]. Our expression analysis found class I gene either in rice or Poplar mainly expressed in pre-emergence inflorescence, post-emergence inflorescence, pistil and 25 DAP embryo (Table 2), showed similar expression patterns as in *Arabidopsis*.

STM (AT1G62360) was classified into class IB in this study and expressed during early embryogenesis, its expression marked the entire SAM (Long *et al.* 1996) and was proposed to be essential for SAM formation and maintenance [41] [42] because STM inhibited the cellular differentiation normally associated with organogenesis and permitted the WUS-CLAVATA feedback loop to maintain the central stem cells [41] [43]. As two Poplar genes, POPTR_0004s00650 and POPRT_0011s01600 exactly fell into STM clade, they had similar characters

as STM in phylogenic, protein structure and expression patterns, they probably functioned as the same as STM in *Arabidopsis* and it was worth to give a further investigation for their roles in cellular differentiation and interactions with other genes.

Two of three rice class IB KNOX genes, LOC_Os03g51690 and LOC_Os07g03770 had highest expression in pre-emergence inflorescence (Table 2) and fell into the same clade to KNAT1. Expression of *Oskn*3 marked the boundaries of different embryonic organs following SAM formation (Postma-Haarsma *et al.* 1999). The expression of KNAT1 covered the embryonic SAM, post-embryonic development SAM and the boundary of the inflorescence SAM (Venglat *et al.* 2002).LOC_Os03g51690 might also take part in the inflorescence architecture just like LOC_Os07g03770 and KNAT1.

KNAT2 was expressed during embryogenesis and marked the base of the SAM [15] [16]. Its' promoter had been reported to be active in root tissue [17]. KNAT6 was expressed in the embryonic SAM, the SAM boundaries [15] and the phloem tissue of roots [18]. The gene POPTR_0008s19300 had strongest expression in nodes and might function in phloem tissue just like KNAT6 as they both fell into the same clade and had the similar expression profiles.

The 4 conserved domains of KNOX protein showed different molecular functions. KNOX1 played a role in suppressing target gene expression,KNOX2 was thought to be necessary for homo-dimerisation, ELK domain was required for the nuclear localization of these proteins, and Homeobox KN domain was a homeobox transcription factor conserved from fungi to human and plants [3] [4] [5]. Poplar gene POPTR_0012s04040, rice genesLOC_Os03g47042 and LOC_Os03g47022 had only the MEINOX domain as the same as KNATM (Figure 2), which indicated the functions owing to ELK and Homeobox-KN4 might be lost in these 3 genes. However, all of them were found to be expressed in nucleus (Table 2 and Table 3) and showed transcription activities by GO analysis (Figure 4 and Figure 5). Though strong functional relationship existed between the homeodomain (HD) and the MEINOX domain, some observations indicated that the MEINOX domain can also work in a homeodomain-independent fashion [44]. Class III members KNATM was found to be participated in transcriptional regulation in a homeodomain-independent fashion [7]. That indicated the KNATM homologues in rice and Poplar might function also independent of Homeoboxdomain. However, the BP-interacting domain reported by Magnani and Kake [7] was not conserved in rice homologues (Figure 3). The mechanisms of KNATM–BP interaction in rice KNATM homologues need further investigations.

Recent evidence suggested that auxin may play a major role in down-regulating KNOX genes during organ emergence [45]. And several clues pointed to a hormone-dependent down-regulation of KNOX genes in the incipient primordium [46] [47], but the identity of the genetic factors responsible for this control is still unknown.

All of these hinted the known and new KNOXs identified in this study may

regulate expression of their target genes to control cell differentiation and development in rice and Poplar. Though further experimental studies still need to be conducted to unravel their biological roles, the genome-wide analysis of KNOXs in rice and Poplar will help to discovery new KNOXs gene and provide a valuable resource for further functional analysis.

Acknowledgements

Conflicts of Interest

The authors declare no competing or financial interests.

References

[1] Kerstetter, R., Vollbrecht, E., Lowe, B., Veit, B., Yamaguchi, J., *et al.* (1994) Sequence-Analysis and Expression Patterns Divide the Maize Knotted1-Like Homeobox Genes into 2 Classes. *Plant Cell*, **6**, 1877-1887. https://doi.org/10.1105/tpc.6.12.1877

[2] Mukherjee, K., Brocchieri, L. and Burglin, T.R. (2009) A Comprehensive Classification and Evolutionary Analysis of Plant Homeobox Genes. *Molecular Biology and Evolution*, **26**, 2775-2794. https://doi.org/10.1093/molbev/msp201

[3] Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., *et al.* (2010) The Pfam Protein Families Database. *Nucleic Acids Research*, **38**, D211-D222. https://doi.org/10.1093/nar/gkp985

[4] Reiser, L., Sanchez-Baracaldo, P. and Hake, S. (2000) Knots in the Family Tree: Evolutionary Relationships and Functions of Knox Homeobox Genes. *Plant Molecular Biology*, **42**, 151-166. https://doi.org/10.1023/A:1006384122567

[5] Hamant, O. and Pautot, V. (2010) Plant Development: A TALE Story. *Comptes Rendus Biologies*, **333**, 371-381. https://doi.org/10.1016/j.crvi.2010.01.015

[6] Burglin, T.R. (1997) Analysis of TALE Superclass Homeobox Genes (MEIS, PBC, KNOX, Iroquois, TGIF) Reveals a Novel Domain Conserved between Plants and Animals. *Nucleic Acids Research*, **25**, 4173-4180. https://doi.org/10.1093/nar/25.21.4173

[7] Magnani, E. and Hake, S. (2008) KNOX Lost the OX: The Arabidopsis KNATM Gene Defines a Novel Class of KNOX Transcriptional Regulators Missing the Homeodomain. *Plant Cell*, **20**, 875-887. https://doi.org/10.1105/tpc.108.058495

[8] Hake, S., Smith, H.M.S., Holtan, H., Magnani, E., Mele, G., *et al.* (2004) The Role of Knox Genes in Plant Development. *Annual Review of Cell and Developmental Biology*, **20**, 125-151. https://doi.org/10.1146/annurev.cellbio.20.031803.093824

[9] Hay, A. and Tsiantis, M. (2010) KNOX Genes: Versatile Regulators of Plant Development and Diversity. *Development*, **137**, 3153-3165. https://doi.org/10.1242/dev.030049

[10] Long, J.A., Moan, E.I., Medford, J.I. and Barton, M.K. (1996) A Member of the KNOTTED Class of Homeodomain Proteins Encoded by the STM Gene of Arabi-

dopsis. *Nature*, **379**, 66-69. https://doi.org/10.1038/379066a0

[11] Truernit, E., Siemering, K.R., Hodge, S., Grbic, V. and Haseloff, J. (2006) A Map of KNAT Gene Expression in the Arabidopsis Root. *Plant Molecular Biology*, **60**, 1-20. https://doi.org/10.1007/s11103-005-1673-9

[12] Byrne, M.E., Simorowski, J. and Martienssen, R.A. (2002) ASYMMETRIC LEAVES1 Reveals Knox Gene Redundancy in Arabidopsis. *Development*, **129**, 1957-1965. http://dev.biologists.org/content/129/8/1957

[13] Douglas, S.J., Chuck, G., Dengler, R.E., Pelecanda, L. and Riggs, C.D. (2002) KNAT1 and ERECTA Regulate Inflorescence Architecture in Arabidopsis. *Plant Cell*, **14**, 547-558. https://doi.org/10.1105/tpc.010391

[14] Venglat, S.P., Dumonceaux, T., Rozwadowski, K., Parnell, L., Babic, V., *et al.* (2002) The Homeobox Gene BREVIPEDICELLUS Is a Key Regulator of Inflorescence Architecture in Arabidopsis. *Proceedings of the National Academy of Sciences USA*, **99**, 4730-4735. https://doi.org/10.1073/pnas.072626099

[15] Belles-Boix, E., Hamant, O., Witiak, S.M., Morin, H., Traas, J., *et al.* (2006) KNAT6: an Arabidopsis Homeobox Gene Involved in Meristem Activity and Organ Separation. *Plant Cell*, **18**, 1900-1907. https://doi.org/10.1105/tpc.106.041988

[16] Dockx, J., Quaedvlieg, N., Keultjes, G., Kock, P., Weisbeek, P., *et al.* (1995) The Homeobox Gene Atk1 of Arabidopsis-Thaliana Is Expressed in the Shoot Apex of the Seedling and in Flowers and Inflorescence Stems of Mature Plants. *Plant Molecular Biology*, **28**, 723-737. https://doi.org/10.1007/BF00021196

[17] Hamant, O., Nogue, F., Belles-Boix, E., Jublot, D., Grandjean, O., *et al.* (2002) The KNAT2 Homeodomain Protein Interacts with Ethylene and Cytokinin Signaling. *Plant Physiology*, **130**, 657-665. https://doi.org/10.1104/pp.004564

[18] Dean, G., Casson, S. and Lindsey, K. (2004) KNAT6 Gene of Arabidopsis Is Expressed in Roots and Is Required for Correct Lateral Root Formation. *Plant Molecular Biology*, **54**, 71-84. https://doi.org/10.1023/B:PLAN.0000028772.22892.2d

[19] Serikawa, K.A., MartinezLaborda, A. and Zambryski, P. (1996) Three Knotted1-Like Homeobox Genes in Arabidopsis. *Plant Molecular Biology*, **32**, 673-683. https://doi.org/10.1007/BF00020208

[20] Birnbaum, K., Shasha, D.E., Wang, J.Y., Jung, J.W., Lambert, G.M., *et al.* (2003) A Gene Expression Map of the Arabidopsis Root. *Science*, **302**, 1956-1960. https://doi.org/10.1126/science.1090022

[21] Li, E.Y., Bhargava, A., Qiang, W.Y., Friedmann, M.C., Forneris, N., *et al.* (2012) The Class II KNOX Gene KNAT7 Negatively Regulates Secondary Wall Formation in Arabidopsis and Is Functionally Conserved in Populus. *New Phytologist*, **194**, 102-115. https://doi.org/10.1111/j.1469-8137.2011.04016.x

[22] Groover, A.T., Mansfield, S.D., DiFazio, S.P., Dupper, G., Fontana, J.R., *et al.* (2006) The Populus Homeobox Gene ARBORKNOX1 Reveals Overlapping Mechanisms Regulating the Shoot Apical Meristem and the Vascular Cambium. *Plant Molecular Biology*, **61**, 917-932. https://doi.org/10.1007/s11103-006-0059-y

[23] Du, J., Mansfield, S.D. and Groover, A.T. (2009) The Populus Homeobox Gene ARBORKNOX2 Regulates Cell Differentiation during Secondary Growth. *Plant Journal*, **60**, 1000-1014. https://doi.org/10.1111/j.1365-313X.2009.04017.x

[24] Postma-Haarsma, A.D., Verwoert, II.G.S., Stronk, O.P., Koster, J., Lamers, G.E.M., *et al.* (1999) Characterization of the KNOX Class Homeobox Genes Oskn2 and Oskn3 Identified in a Collection of cDNA Libraries Covering the Early Stages of Rice Embryogenesis. *Plant Molecular Biology*, **39**, 257-271. https://doi.org/10.1023/A:1006153506868

[25] Sentoku, N., Sato, Y., Kurata, N., Ito, Y., Kitano, H., *et al.* (1999) Regional Expression of the Rice KN1-Type Homeobox Gene Family during Embryo, Shoot, and Flower Development. *Plant Cell*, **11**, 1651-1663.
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC144314/
https://doi.org/10.1105/tpc.11.9.1651

[26] Sato, Y., Aoki, Y. and Matsuoka, M. (2002) A Loss-of-Function Mutation in the Rice KNOX Type Homeobox Gene, OSH3. *Plant and Cell Physiology*, **43**, 44-51.
https://doi.org/10.1093/pcp/pcf003

[27] Ito, Y., Hirochika, H. and Kurata, N. (2002) Organ-Specific Alternative Transcripts of KNOX Family Class 2 Homeobox Genes of Rice. *Gene*, **288**, 41-47.
https://doi.org/10.1016/S0378-1119(02)00460-2

[28] Kuijt, S.J.H., Lamers, G.E.M., Rueb, S., Scarpella, E., Ouwerkerk, P.B.F., *et al.* (2004) Different Subcellular Localization and Trafficking Properties of KNOX Class 1 Homeodomain Proteins from Rice. *Plant Molecular Biology*, **55**, 781-796.
https://doi.org/10.1007/s11103-005-1967-y

[29] Tsuda, K., Ito, Y., Yamaki, S., Miyao, A., Hirochika, H., *et al.* (2009) Isolation and Mapping of Three Rice Mutants That Showed Ectopic Expression of KNOX Genes in Leaves. *Plant Science*, **177**, 131-135. https://doi.org/10.1016/j.plantsci.2009.04.007

[30] Altschul, S., Madden, T., Schaffer, A., Zhang, J.H., Zhang, Z., *et al.* (1998) Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Faseb Journal*, **12**, A1326-A1326.

[31] Guo, A., Zhu, Q., Chen, X. and Luo, J. (2007) GSDS: A Gene Structure Display Server. *Yi Chuan*, **29**, 1023-1026. https://doi.org/10.1360/yc-007-1023

[32] Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., *et al.* (2009) MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic Acids Research*, **37**, W202-W208. https://doi.org/10.1093/nar/gkp335

[33] Marchler-Bauer, A., Lu, S.N., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., *et al.* (2011) CDD: A Conserved Domain Database for the Functional Annotation of Proteins. *Nucleic Acids Research*, **39**, D225-D229. https://doi.org/10.1093/nar/gkq1189

[34] Kelley, L.A. and Sternberg, M.J.E. (2009) Protein Structure Prediction on the Web: A Case Study using the Phyre Server. *Nature Protocols*, **4**, 363-371.
https://doi.org/10.1038/nprot.2009.2

[35] Du, Z., Zhou, X., Ling, Y., Zhang, Z.H. and Su, Z. (2010) agriGO: A GO Analysis Toolkit for the Agricultural Community. *Nucleic Acids Research*, **38**, W64-W70.
https://doi.org/10.1093/nar/gkq310

[36] Kaundal, R., Saini, R. and Zhao, P.X. (2010) Combining Machine Learning and Homology-Based Approaches to Accurately Predict Subcellular Localization in Arabidopsis. *Plant Physiology*, **154**, 36-54. https://doi.org/10.1104/pp.110.156851

[37] Sjodin, A., Street, N.R., Sandberg, G., Gustafsson, P. and Jansson, S. (2009) The Populus Genome Integrative Explorer (PopGenIE): A New Resource for Exploring the Populus Genome. *New Phytologist*, **182**, 1013-1025.
https://doi.org/10.1111/j.1469-8137.2009.02807.x

[38] Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: Discovering Splice Junctions with RNA-Seq. *Bioinformatics*, **25**, 1105-1111.
https://doi.org/10.1093/bioinformatics/btp120

[39] Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., *et al.* (2010) Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching during Cell Differentiation. *Nature Biotechnology*, **28**, 511-U174. https://doi.org/10.1038/nbt.1621

[40] Atchley, W.R. and Fitch, W.M. (1997) A Natural Classification of the Basic Helix-Loop-Helix Class of Transcription Factors. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 5172-5176.
https://doi.org/10.1073/pnas.94.10.5172

[41] Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F.X., Jurgens, G., *et al.* (2000) The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. *Cell*, **100**, 635-644.
https://doi.org/10.1016/S0092-8674(00)80700-X

[42] Scofield, S., Dewitte, W. and Murray, J.A. (2008) A Model for Arabidopsis Class-1 KNOX Gene Function. *Plant Signaling & Behavior*, **3**, 257-259.
https://doi.org/10.4161/psb.3.4.5194

[43] Brand, U., Fletcher, J.C., Hobe, M., Meyerowitz, E.M. and Simon, R. (2000) Dependence of Stem Cell Fate in Arabidopsis on a Feedback Loop Regulated by CLV3 Activity. *Sciences*, **289**, 617-619. https://doi.org/10.1126/science.289.5479.617

[44] Noro, B., Culi, J., Mckay, D.J., Zhang, W. and Mann, R.S. (2006) Distinct Functions of Homeodomain-Containing and Homeodomain-Less Isoforms Encoded by Homothorax. *Genes & Development*, **20**, 1636-1650.
https://doi.org/10.1101/gad.1412606

[45] Scanlon, M.J. (2003) The Polar Auxin Transport Inhibitor N-1-Naphthylphthalamic Acid Disrupts Leaf Initiation, KNOX Protein Regulation, and Formation of Leaf Margins in Maize. *Plant Physiology*, **133**, 597-605.
https://doi.org/10.1104/pp.103.026880

[46] Sakamoto, T., Kamiya, N., Ueguchi-Tanaka, M., Iwahori, S. and Matsuoka, M. (2001) KNOX Homeodomain Protein Directly Suppresses the Expression of a Gibberellin Biosynthetic Gene in the Tobacco Shoot Apical Meristem. *Genes & Development*, **15**, 581-590. https://doi.org/10.1101/gad.867901

[47] Hay, A., Kaur, H., Phillips, A., Hedden, P., Hake, S., *et al.* (2002) The Gibberellin Pathway Mediates KNOTTED1-Type Homeobox Function in Plants with Different Body Plans. *Current Biology*, **12**, 1557-1565.
https://doi.org/10.1016/S0960-9822(02)01125-9

Appendix

All supplementary materials can be found in this link:

https://pan.baidu.com/s/1QOnjLDMr1CMbkKTk8e6RxQ!