Molecular Plant Breeding 2021, Vol.12, No.19, 1-11
http://genbreedpublisher.com/index.php/mpb

Research Article Open Access

Identification and Bioinformatics Analysis of *KNOX* Gene Family in Wheat (*Triticum aestivum* L.)

Song Ningning, Liang Huihui, An Yiwei, Bai Shenglong, Ma Feifei, Zhang Zhen, Li Hao, Zhou Yun, Guo Guanghui ✉, Song Chunpeng

Henan University School of Life Sciences, State Key Laboratory of Crop Stress Adaptation and Improvement, KaiFeng, 475004, P.R. China

✉ Corresponding author email: guanghuiguo2008@163.com

Molecular Plant Breeding, 2021, Vol.12, No.19 doi: 10.5376/mpb.2021.12.0019

Received: 02 Jun., 2021

Accepted: 13 Jun., 2021

Published: 20 Jun., 2021

Copyright © 2021 Song et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Song N.N., Liang HH., An Y.W., Bai S.L., Ma F.F., Zhang Z., Li H., Zhou Y., Guo G.H., and Song C.P., 2021, Identification and bioinformatics analysis of *KNOX* gene family in wheat (*Triticum aestivum* L.), Molecular Plant Breeding, 12(19): 1-11 (doi: 10.5376/mpb.2021.12.0019)

Abstract KNOTTED-like homeodomain (*KNOX*) gene family is a transcription factor encoding homeobox protein, which plays an important role in plant growth and morphogenesis. However, little information is available on the *KNOX* gene family in wheat (*Triticum aestivum* L.). In this study, 36 *KNOX* genes with KNOX1 or KNOX2 domain distributed on 18 chromosomes were identified from wheat genome using bioinformatics methods. The phylogenetic evolution, gene structure, protein domains, cis-acting elements and gene expression patterns were analyzed in the present study. Based on the phylogenetic tree, the 36 *TaKNOX* genes were divided into two major subclasses, Class I and Class II, and the two subclasses were further divided into five evolutionary branches. Most *TaKNOX* genes contain four typical conserved protein domains: KNOX1, KNOX2, ELK and HOX. Some cis-acting elements are associated with hormonal, plant development and stress in *TaKNOX* promoters. The analysis result of transcriptome data from wheat different tissue showed that Class I *KNOX* genes had obvious tissue specificity, while Class II *KNOX* genes expressed widely in different wheat tissues. The study results provide important information for future analysis of the regulation and functions of the *TaKNOX* gene family.

Keywords Wheat; *KNOX* gene family; Gene structure; Gene expression

KNOX (KNOTTED1-like homeobox) transcription factor family is a subfamily of the homeobox gene TALE (Three amino acid loop extension) protein superfamily, which is widely involved in the regulation of plant growth and development (Hamant and Pautot, 2010). Most KNOX proteins contain four typical conserved domains: KNOX1, KNOX2, ELK and HOX (Magnani and Hake, 2008; Gao et al., 2015).

KNOX gene family is a conservative gene family in the plant kingdom, which is widespread from lower plant phycophyte and bryophyte to higher plant spermatophyte (Gao et al. 2015). The first *KNOX* gene to be identified in plants was Knotted1 (*Kn1*) in maize (Vollbrecht et al., 1991). Following this discovery, the *KNOX* gene was identified in more and more plants (Gao et al., 2015), such as Arabidopsis, rice, poplar, cotton, apple and so on, and the functions of some *KNOX* genes have been deeply studied (Mukherjee et al., 2009; Xiong et al., 2018; Ma et al., 2019; Jia et al., 2020). However, little information is available on the *KNOX* gene family in the important food crop wheat (*Triticum aestivum* L.). At present, only a report about Takumi et al. (2000) cloned three *KNOX* genes from wheat with the help of maize *Kn1* gene sequence.

Based on the structural characteristics, phylogenetic tree and expression pattern of *KNOX* genes, the *KNOX* gene family can be divided into three types: Class I, Class II, and Class KNATM, which is endemic to dicotyledons (Magnani and Hake, 2008; Gao et al., 2015; Xiong et al., 2018). The expression patterns and biological functions of KNOX genes of different classes have obvious differentiation. The expression of Class I genes is concentrated, mainly in SAM (Shoot apical meristem), and they play an important role in the differentiation and maintenance of meristem (Tsuda et al., 2011; Gao et al., 2015; Su et al., 2020). For example, the *STM* gene in Arabidopsis Class I plays an important role in the establishment of SAM in Arabidopsis embryos, and the seeds of Arabidopsis *stm* mutants can only produce cotyledons but not new leaves because of the lack of SAM. The *KNOX* gene in Class I is closely related to callus differentiation in the process of transgenic. The calli of the double deletion mutants of

osh1 and *osh15* genes in rice Class I can only form leaf-like structures, but no bud formation can be seen (Tsuda et al., 2011). When the *KN1* gene of maize Class I was transferred into tobacco, the rate of seedling transformation was twice as high as that of the control (Xu et al., 2009). At the same time, Class I genes also play an important role in the control of leaf shape, internode elongation, hormone balance and the establishment of inflorescence structure, and can be used as transcriptional activators or suppressors (Hay and Tsiantis, 2010; Tsuda and Hake, 2015). The *KNOX* subfamily genes of Class II are generally widely expressed in various tissues and organs of plants, but there are relatively few reports about the function of Class II *KNOX* genes in plants, which are mainly related to the formation of the secondary cell wall. For example, the synergistic effect of *KNAT7* and *KNAT3* affects the deposition of the secondary cell wall and improves the mechanical support of the Arabidopsis stem (Wang et al., 2020). Class KNATM is a kind of *KNOX* gene subfamily peculiar to dicotyledons. It is found that *KNATM* gene is involved in regulating leaf polarity and leaf shape in *Arabidopsis thaliana* (Magnani and Hake, 2008).

As one of the most important food crops in the world, wheat is the staple food for about 35%~40% of the world's population (He et al., 2018). However, due to the complexity of wheat genome, there are few reports on the cloning and research of the *TaKNOX* gene. The continuous improvement of wheat genome information (Iwgsc, 2018), it brings an opportunity to identify and analyze *TaKNOX* gene at the whole genome level. In the current study, *TaKNOX* gene was identified by using wheat genome information, combined with the characteristics of *KNOX* gene family, and the basic physical and chemical properties, chromosome distribution, gene replication, evolutionary relationship, gene structure, cis-acting elements and *KNOX* gene expression patterns in different tissues were analyzed, which laid a foundation for further study on the function and regulation mechanism of *TaKNOX* gene.

1 Results and Analysis

1.1 Identification and nomenclature of *TaKNOX* gene

Based on the amino acid sequence of Arabidopsis and rice KNOX protein, BLAST search, screening and verification were carried out in the wheat database, and a total of 36 *TaKNOX* genes were identified, which were named *TaKNOX1*~ *TaKNOX36* (Table 1) according to the chromosome location of the gene. Among the proteins encoded by the *TaKNOX* gene family, the number of amino acids ranges from 153~389. TaKNOX protein MW range from 16.54~42.48 kDa, and the PI is 5.15~9.11. Except for TaKNOX13, TaKNOX15, TaKNOX18 and TaKNOX34, the PI of other TaKNOX proteins is less than 7, indicating that most TaKNOX proteins are acidic. Subcellular localization prediction analysis showed that most of *KNOX* genes were located in the nucleus, which was consistent with the transcription factor characteristics of *KNOX* gene.

1.2 Multiple alignments of TaKNOX proteins

Multiple sequence alignment showed that TaKNOX protein sequence contains four relatively conserved regions: KNOX1, KNOX2, ELK and HOX (Figure 1A). KNOX1 and KNOX2 domains are located at the N-terminal, while ELK and HOX are located at the C-terminal. Among them, HOX domain is the most conservative, showing the typical structural characteristics of the TALE homeobox protein superfamily, and there are three additional amino acid sequences between the first and second helix (P-Y-P). KNOX1, KNOX2 and ELK also showed effective conservatism. For example, KNOX2 contains a highly conserved E-L-D amino acid sequence, and the ELK contains a highly conserved E-L-K amino acid sequence (Figure 1B).

1.3 Chromosome mapping and gene duplication analysis of *TaKNOX* gene

KNOX gene was distributed on all chromosomes except chromosome 3 in wheat, among which there was one *KNOX* gene on chromosomes 2, 6 and 7 (Figure 2). *KNOX* gene is the most common on chromosome 4A, with 6 *KNOX* genes. At the same time, most *KNOX* genes are distributed at the ends of chromosomes. Segmental duplication and tandem repeat are the main mechanisms of gene family expansion (Cannon et al., 2004). Collinear analysis of *TaKNOX* gene showed that it contained five pairs of tandemly repeated genes: *TaKNOX1* and *TaKNOX2*, *TaKNOX3* and *TaKNOX4*, *TaKNOX6* and *TaKNOX7*, *TaKNOX14* and *TaKNOX15*, *TaKNOX22* and *TaKNOX23* (In Figure 2, the red font is tandem repeated gene). In addition, 23 pairs of *TaKNOX* genes were found

fragment repeat block, including most of the *TaKNOX* genes with partial homology (In Figure 2, the blue line represents the *KNOX* gene in the fragment repeat block). The results showed that there were tandem repeats and segmental duplication in the process of polyploidy in the expansion of *TaKNOX* gene.

Table1 Physicochemical properties of *TaKNOX* gene family

Gene name	Gene ID	Amino size (aa)	MW (Da)	PI	Subcellular localization
TaKNOX1	TraesCS1A02G072700	308	33778.98	5.66	Nucl;cyto_nucl
TaKNOX2	TraesCS1A02G072800	307	33144.26	5.40	Nucl
TaKNOX3	TraesCS1B02G091600	308	33523.78	5.80	Nucl
TaKNOX4	TraesCS1B02G091700	307	33202.34	5.40	Nucl
TaKNOX5	TraesCS1B02G135600	282	31432.56	5.88	Vacu; chlo; nucl; extr; golg
TaKNOX6	TraesCS1D02G075600	307	33652.81	5.66	Nucl
TaKNOX7	TraesCS1D02G075700	307	33174.29	5.40	Nucl
TaKNOX8	TraesCS2A02G267400	348	38522.37	6.09	Nucl
TaKNOX9	TraesCS2B02G268200	348	38560.47	6.09	Nucl
TaKNOX10	TraesCS2D02G256400	348	38463.30	5.98	Nucl
TaKNOX11	TraesCS4A02G251800	167	18107.19	5.15	Mito; Chlo;Nucl
TaKNOX12	TraesCS4A02G256700	363	40079.08	6.28	Nucl
TaKNOX13	TraesCS4A02G291900	330	36155.39	8.61	Nucl
TaKNOX14	TraesCS4A02G292000	153	16539.84	6.10	Chlo; Mito; nucl; Nucl_plas
TaKNOX15	TraesCS4A02G292100	331	36332.65	8.61	Nucl
TaKNOX16	TraesCS4A02G292200	320	35589.70	5.50	Nucl
TaKNOX17	TraesCS4B02G021900	325	36308.44	5.80	Nucl
TaKNOX18	TraesCS4B02G022000	301	33269.52	9.11	Nucl
TaKNOX19	TraesCS4B02G057900	364	40193.18	6.28	Nucl
TaKNOX20	TraesCS4B02G346100	313	34437.99	5.76	Nucl
TaKNOX21	TraesCS4D02G019600	341	38120.78	6.23	Nucl
TaKNOX22	TraesCS4D02G019800	323	35512.79	6.32	Nucl
TaKNOX23	TraesCS4D02G019900	327	35765.91	6.08	Nucl
TaKNOX24	TraesCS4D02G058000	363	40065.05	6.28	Nucl
TaKNOX25	TraesCS4D02G341000	312	34496.07	5.86	Nucl
TaKNOX26	TraesCS5A02G405900	334	36772.57	5.97	Nucl;cyto_nucl
TaKNOX27	TraesCS5A02G515000	313	34437.99	5.76	Nucl
TaKNOX28	TraesCS5B02G410600	317	35314.09	5.96	Nucl
TaKNOX29	TraesCS5B02G454600	177	20218.73	5.30	Nucl; Cyto; Chlo; Plas
TaKNOX30	TraesCS5D02G415900	340	37367.24	5.97	Nucl
TaKNOX31	TraesCS6A02G145500	307	33472.45	5.72	Nucl
TaKNOX32	TraesCS6B02G173900	299	32459.32	5.69	Nucl
TaKNOX33	TraesCS6D02G134800	208	23540.29	6.10	Nucl; Cyto_nucl
TaKNOX34	TraesCS7A02G511800	389	42477.98	7.66	Nucl
TaKNOX35	TraesCS7B02G423100	293	32422.46	5.90	Nucl
TaKNOX36	TraesCS7D02G501000	310	33985.99	5.69	Nucl

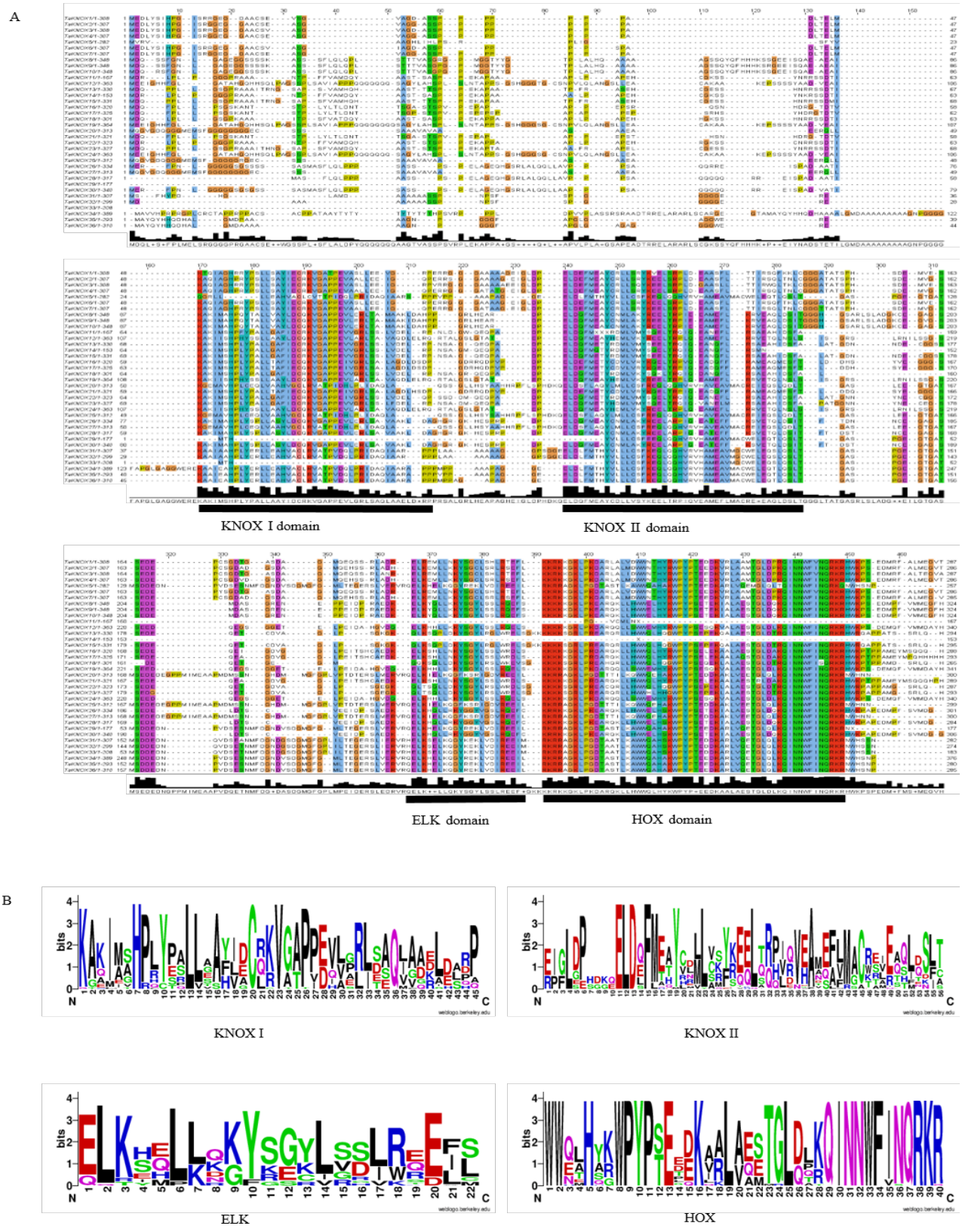

Figure 1 Multiple alignments of KNOX proteins in wheat

Note: The black horizontal line in Figure A indicates the four conserved domains of KNOXI, KNOXII, ELK and HOX; Figure B shows the distribution and composition of protein characteristic motifs of each conserved domain; Letters stand for the amino acids; The height of the letters show the conservation level

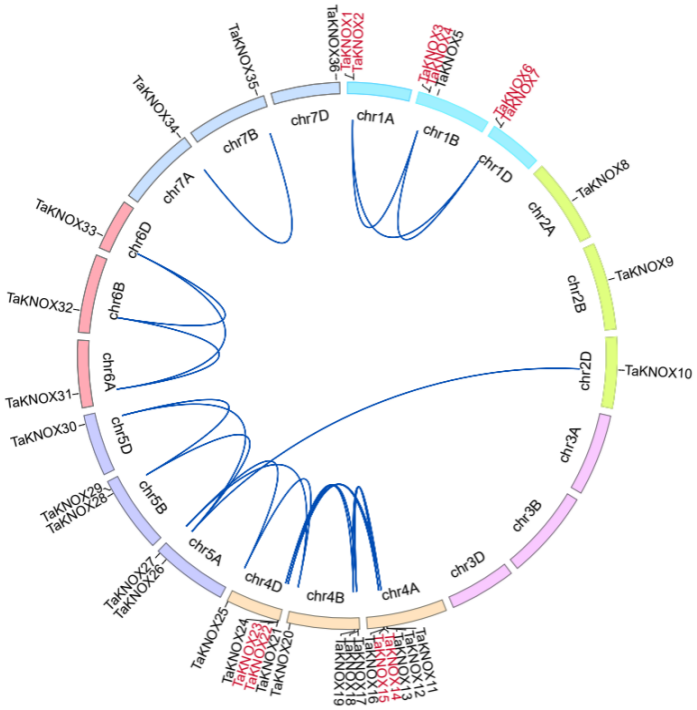

Figure 2 Chromosome mapping and gene duplication of *KNOX* genes in the wheat genome

Note: The blue line represents the *TaKNOX* gene pair in the fragment repeat block; The red font is tandem repeated *TaKNOX* gene

1.4 Phylogenetic analysis of *TaKNOX* gene

KNOX protein sequences of Arabidopsis, rice and the KNOX protein sequences of wheat were used to construct the KNOX phylogenetic tree using MEGA X software (Figure 3). Based on the phylogenetic tree, KNOX gene was divided into three categories: Class I, Class II and Class KNATM. Class I contains 38 genes, including 4 Arabidopsis genes, 9 rice genes, and 25 wheat genes. Class I contains 19 genes, including 4 Arabidopsis genes, 4 rice genes, and 11 wheat genes. Class KNATM contains only one member (*KNATM*) of Arabidopsis, which is unique to dicotyledons (Magnani and Hake, 2008).

Class I was further divided into three subclasses: ClassIA, ClassIB, and ClassIC. ClassII contains ClassIIA and ClassIIB (Figure 3). *TaKNOX19*, *TaKNOX24* and *TaKNOX26* of ClassIA were clustered with *AtKNAT1* and *AtSTM* of *Arabidopsis thaliana* and *LOC_Os03g51690*(*OSH1*) of rice. *TaKNOX1*, *TaKNOX2*, *TaKNOX3*, *TaKNOX4*, *TaKNOX6*, and *TaKNOX7* of ClassIB were clustered with *AtKNAT2* and *AtKNAT6* of *Arabidopsis thaliana*. *TaKNOX20*, *TaKNOX25*, and *TaKNOX27* of ClassIIA were clustered with *AtKNAT7* of *Arabidopsis thaliana* (Figure 3). These *TaKNOX* genes may have similar biological functions to Arabidopsis and rice KNOX genes which are closely related to their phylogeny. At the same time, some of the *KNOX* homologous genes located in A, B and D subgenomes are closely related, and most of them are clustered together, such as *TaKNOX1*, *TaKNOX3*, *TaKNOX6*, *TaKNOX2*, *TaKNOX4*, *TaKNOX7*, *TaKNOX8*, *TaKNOX9*, *TaKNOX10*, *TaKNOX12*, *TaKNOX19*, *TaKNOX24*, *TaKNOX26*, *TaKNOX28*, *TaKNOX30* etc.

1.5 The gene structure and protein conserved domain analysis of *TaKNOX* gene

To further explore the characteristics of *TaKNOX* gene family, based on an intraspecific phylogenetic tree (Figure 4A), we analyzed the gene structure and protein conserved domain distribution of *TaKNOX* gene (Figure 4). The results showed that the TaKNOX family was mainly composed of 3～6 exons and 4～5 introns. The number of exons in ClassI ranges from 4～6, while almost all genes in ClassII contain 5 exons (Figure 4B). It is worth noting that all ClassI genes, except *TaKNOX11* and *TaKNOX14*, contain a long intron (Figure 4B), which is consistent with the structural characteristics of the ClassI *KNOX* gene family (Morimoto et al., 2005). Most TaKNOX proteins contain conserved protein domains of KNOX1, KNOX2, ELK and HOX, and are arranged in turn on the protein sequence. Some TaKNOX proteins lack two conserved domains, for example, TaKNOX11 and TaKNOX14 lack ELK and HOX conserved domains, TaKNOX29 and TaKNOX33 lack KNOX1 domain. But all

TaKNOX proteins contain KNOX2 (Figure 4C). Meanwhile, based on the perspective of evolutionary relationships, the *TaKNOX* genes of the same subfamily are also conservative in gene structure and protein domain distribution.

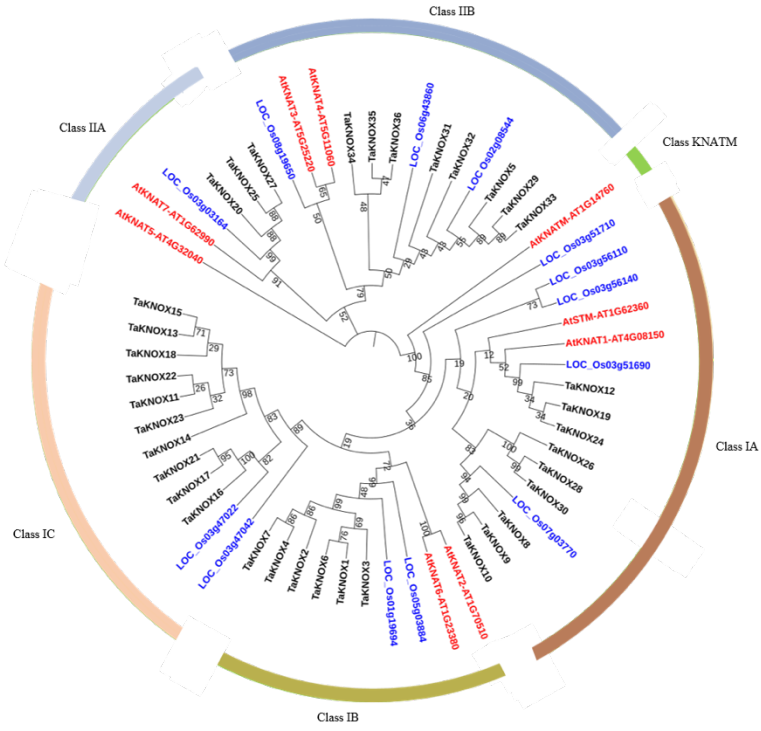

Figure 3 The phylogenetic tree of *KNOX* gene family in wheat, Arabidopsis and rice

Note: *KNOX* genes from wheat, Arabidopsis and rice are represented in black, red and blue fonts, respectively

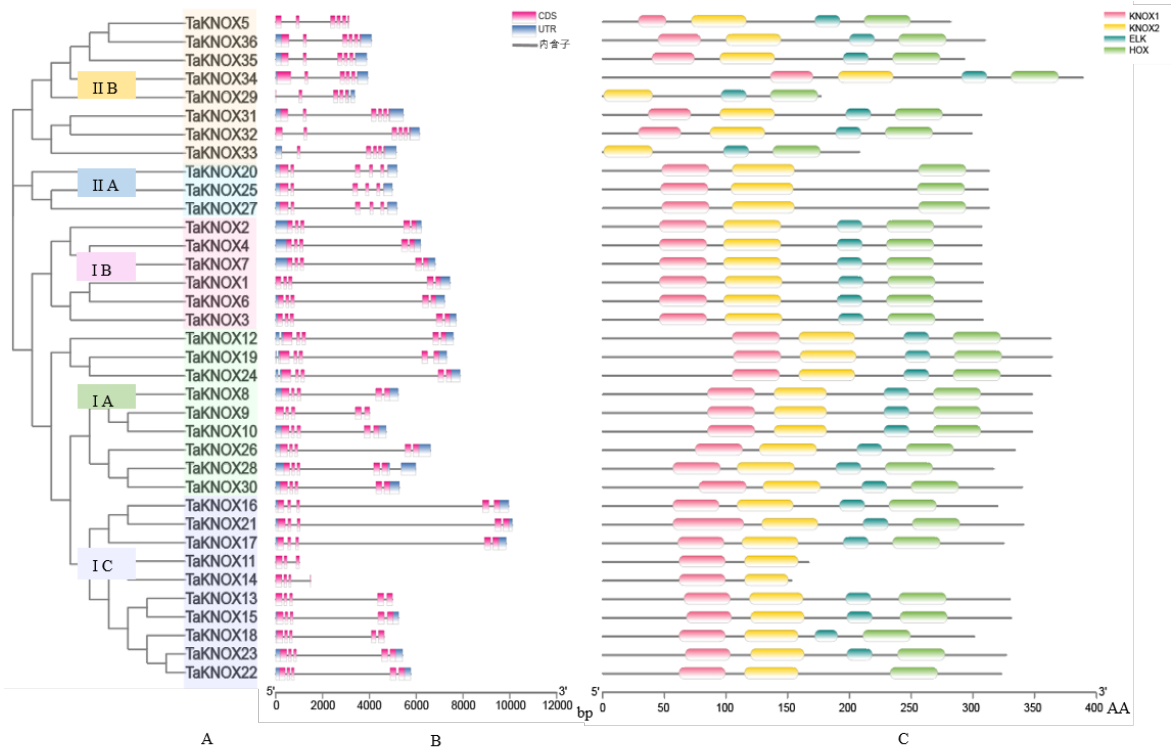

Figure 4 The gene structure and protein conserved domain analysis of *KNOX* gene family in wheat

Note: A: The phylogenetic tree of *KNOX* gene family in wheat; B: The gene structure composition of *KNOX* gene family in wheat; C: The conserved domain composition of *KNOX* gene family in wheat

1.6 Analysis of cis-acting elements on the promoter of *TaKNOX* gene

In order to analyze the potential expression regulation mechanism of *TaKNOX* gene family members, we identified the cis-acting elements of each member. We selected the cis-acting elements related to plant hormones, plant growth, stress response and light response for identifying (Figure 5). The results showed that the promoter regions of all *TaKNOX* genes contained elements related to plant hormones and stress response. For example, the cis-acting elements related to plant hormones are ABA response element ABRE, MeJA response element CGTCA-motif, TGACG-motif, GA response element TATC-box, GARE-motif, P-box and so on. The cis-acting elements related to stress response are low-temperature induction response element LTR, defense and stress related response element TC-rich repeats and so on. In addition, 28 *TaKNOX* genes contain cis-acting elements (O2-site, CAT-box, GCN4-motif, circadian) related to the regulation of plant growth and development. Moreover, 23 *TaKNOX* genes contain elements (CAT-box) related to the expression of plant meristem. These results showed that *TaKNOX* gene may play an important role in wheat growth and development, maintenance of meristem differentiation ability and response to stress (Figure 5).

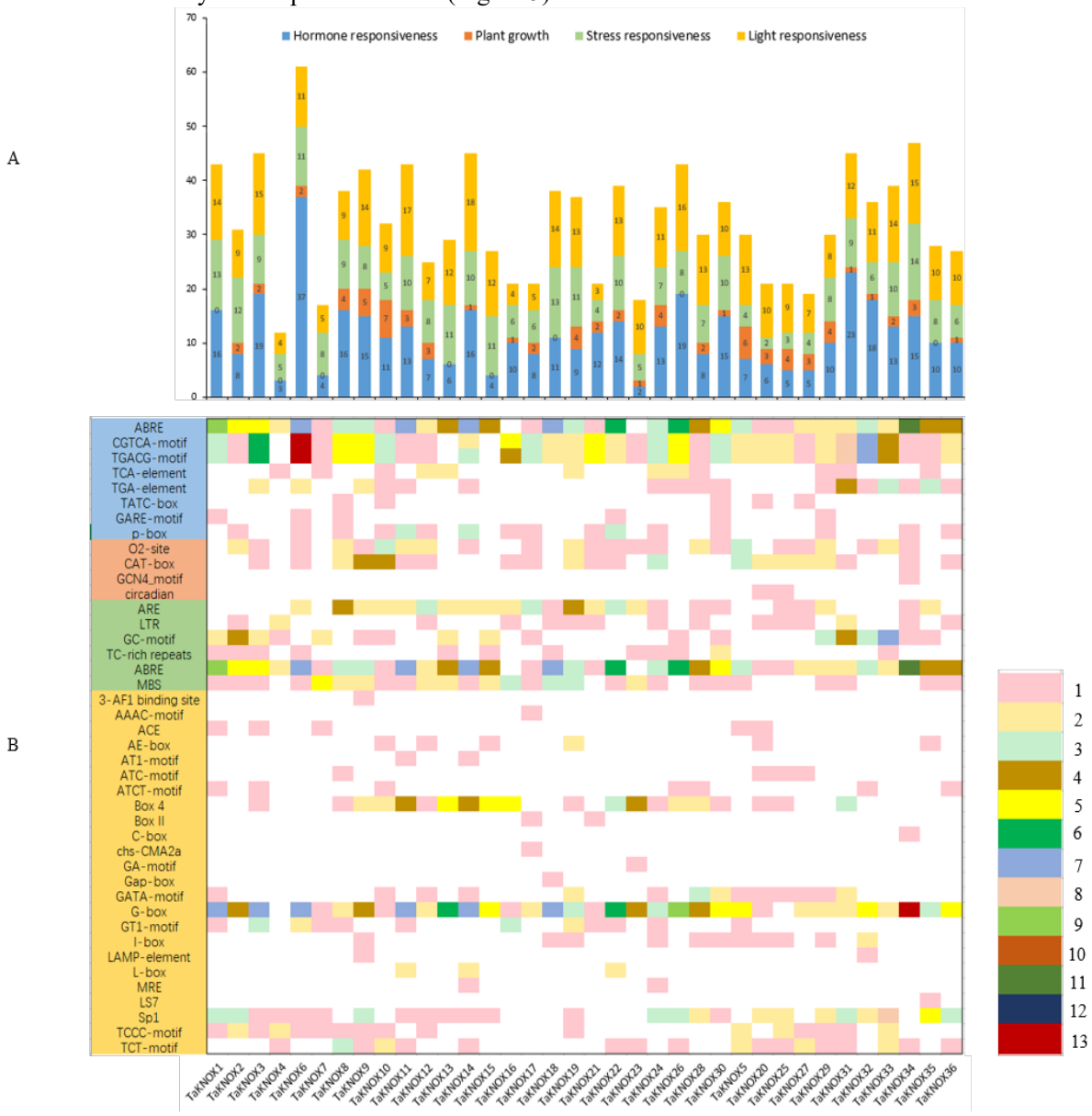

Figure 5 Analysis cis-elements of the *KNOX* genes in wheat

Note: A: Histograms of different colors are used to represent the sum of four kinds of cis-acting elements in *TaKNOX* genes; B: Heat map cis-acting elements of each *TaKNOX* gene. Different colors indicate the number of cis-acting elements, while white indicates that there are no cis-acting elements

1.7 Expression patterns of *TaKNOX* genes in different tissues

In order to explore the tissue-specific expression patterns of *KNOX* gene in wheat at different growth and developmental stages, the expression data of 36 *TaKNOX* genes were analyzed, and the transcriptional levels of wheat roots, stems, leaves, spikes and seeds were studied. It was found that there were significant differences in the expression patterns of wheat *TaKNOX* gene in different tissues or organs (Figure 6). There is a certain correlation between the expression pattern of *TaKNOX* gene and its subclass. In Class I, except *TaKNOX18* was mainly expressed in leaves, other members were mainly expressed in stems and panicles, and *TaKNOX11*, *TaKNOX13*, *TaKNOX14* and *TaKNOX15* were highly expressed in panicles. While in Class II, except *TaKNOX34*, *TaKNOX35* and *TaKNOX36* were highly expressed in leaf and other genes were widely expressed.

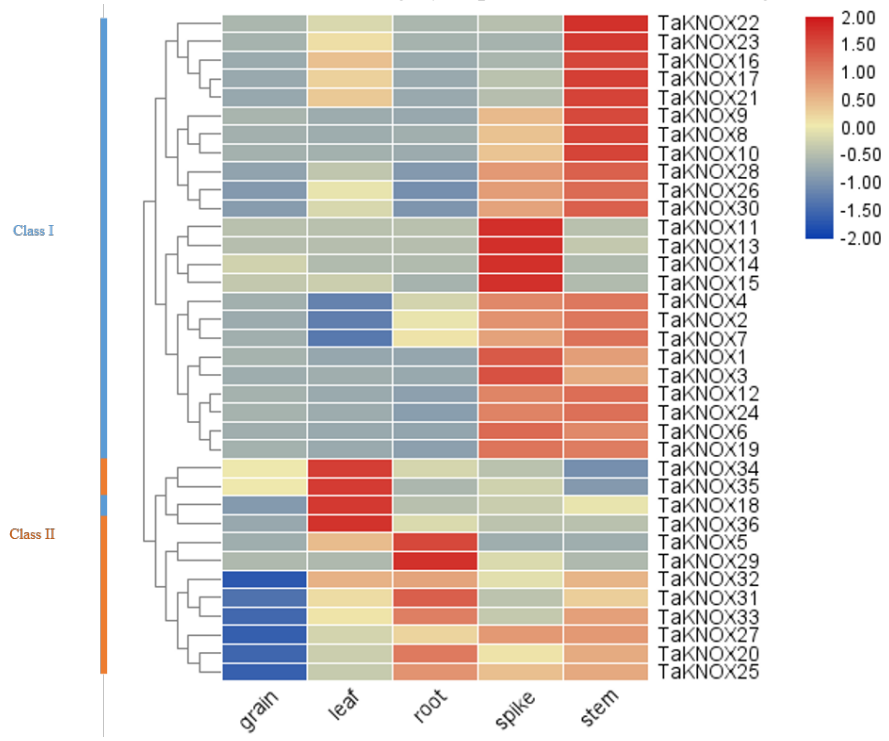

Figure 6 *TaKNOX* gene expression profiles in different tissues

2 Discussion

As an important regulatory factor of plant growth and development, *KNOX* gene lacks identification and functional research in wheat. At present, little information is available on the *KNOX* gene family in wheat (*Triticum aestivum* L.). Only Takumi et al. (2000) cloned three *KNOX* genes (*TaKNOX12*, *TaKNOX19*, and *TaKNOX24* in this study) from wheat with the help of maize *Kn1* gene sequence. Morimoto et al. (2005) analyzed the expression patterns of these three *TaKNOX* genes and the phenotypes of transgenic tobacco. With the continuous improvement of wheat genome information, it brings an opportunity to identify and analyze TaKNOX transcription factors at the whole genome level. In this study, the *KNOX* gene family of wheat was identified at the whole genome level for the first time, and the basic physical and chemical properties, gene structure, evolutionary relationship and expression pattern of *TaKNOX gene* were analyzed, which laid a foundation for future functional study of *KNOX* gene.

In this study, we identified 36 members of the *KNOX* gene family, more than Arabidopsis (9) and rice (13). Among these genes, there are 5 pairs of tandem repeats and 28 pairs of segmental duplication, indicating that the whole genome repetition plays an important role in the amplification of *TaKNOX* gene family, which leads to the larger size of *TaKNOX* gene family than Arabidopsis and rice. According to the protein sequence, evolutionary tree and gene structure characteristics, *TaKNOX* genes were divided into two categories: Class I and Class II. Among them, *TaKNOX1~4*, *TaKNOX6~19*, *TaKNOX21~24*, *TaKNOX26*, *TaKNOX28* and *TaKNOX30* exist in the same evolutionary branch with Class I genes in Arabidopsis and rice, and the remaining *TaKNOX* genes are

clustered into one branch with Class II genes in Arabidopsis and rice. It is worth noting that *TaKNOX12*, *TaKNOX19* and *TaKNOX24* in Class I are most closely related to Arabidopsis *AtSTM* (*AT1G6230*), *AtKNAT1* (*AT4G08150*) and rice *OSH1* (*LOC_Os03g51690*). These three *KNOX* genes may have similar biological functions to *AtSTM*, *AtKNAT1* and *OSH1*, that is, affecting the establishment of SAM in embryos and the formation of embryogenic calli (Hay and Tsiantis, 2010; Tsuda et al., 2011). Therefore, these three *TaKNOX* genes may be used to improve the efficiency of embryogenic callus formation in the process of wheat genetic transformation, to improve the genetic transformation efficiency of wheat.

The results of *TaKNOX* gene transcriptome data showed that the expression level of Class I genes (except for *TaKNOX18* gene) were very low or non-expressed in other tissues, except in panicle and stem. However, Class II genes are expressed in almost all tissues of wheat, which is consistent with the expression patterns of KNOX I and KNOX II genes in other plants, indicating that plant KNOX proteins may be highly conserved in function (Gao et al., 2015). Moreover, the expression of *TaKNOX8*, *TaKNOX9*, *TaKNOX10*, *TaKNOX12*, *TaKNOX19* and *TaKNOX24* of Class I genes in wheat stem is the highest, and the expression is the highest in the early stage of stem development, and gradually decreases with the late development, which may indicate that these genes play an important role in the maintenance and regulation of wheat stem meristem. The function of these genes needs to be further investigated. Therefore, deeply understanding the function of *TaKNOX* gene will lay an important foundation for the study of wheat growth and development, morphogenesis and other processes.

3 Materials and Methods

3.1 Identification of *TaKNOX* gene family

To identify all *KNOX* genes in wheat genome database, the main steps are as follows: (1) Download the genomic data of Chinese Spring from Ensemble Plants database (http://plants.ensembl.org/Triticum_aestivum/Info/Index). (2) Download 9 KNOX protein sequences of Arabidopsis and 13 KNOX protein sequences of rice from TAIR website (https://www.arabidopsis.org/) and Rice Genome Annotation Project website (http://rice.plantbiology.msu.edu/index.shtml), respectively. (3) The KNOX protein sequences of Arabidopsis and rice were used as seed sequences, and the blast command of Tbtools (Chen et al. 2020) was used to search the wheat protein database, with the threshold of e-value < e^{-10}. (4) Submit the search results to the "PFAM" website (http://pfam.xfam.org/) for further verification of the Pfam number (PF03790 or PF03791) of KNOX conserved domain KNOX I or KNOX II (Finn et. al. 2016). Those with KNOX I or KNOX II domains are regarded as members of the *KNOX* gene family. The number of amino acids (AA), molecular weight (MW) and isoelectric point (PI) of each KNOX protein were predicted by ProtParam website (https://web.expasy.org/protparam/), and the subcellular localization of KNOX protein was predicted on WoLF PSORT website (https://wolfpsort.hgc.jp/).

3.2 Protein domain analysis and multiple sequence alignment of members of *TaKNOX* gene family

Analyzed the KNOX protein domain of wheat with the help of CDD website (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and Pfam website (http://pfam.xfam.org/). The multiple sequence alignment of KNOX protein sequence was carried out by using MUSCLE in MEGA X(Kumar et al., 2018), and the visualization was carried out by Jalview software.

3.3 Chromosome mapping and gene replication of members of *TaKNOX* gene family

The chromosome position information of the members of the *KNOX* gene family was obtained according to the wheat genome annotation file. MCScanX software (Wang et al., 2012) was used to calculate and obtain the tandem repeat information of collinear blocks and genes. Circos software (Krzywinski et al., 2009) was used to visualize the chromosome position information, collinear relationship, tandem repeat information of *TaKNOX* gene.

3.4 Phylogenetic analysis of *KNOX* gene family

The members of *KNOX* gene family of Arabidopsis, rice and wheat were used to construct a phylogenetic tree by MEGA X software. The multiple alignments of KNOX protein sequence were carried out by using the MUSCLE alignment method, and Neighbor-joining was selected to construct the phylogenetic tree. Specific parameters are as follows: Bootstrap method with parameters of 1 000, Poisson model.

3.5 Analysis of cis-acting elements in the promoter region of members of *TaKNOX* gene family

The promoter region of 2 000 bp upstream of the CDS sequence of each *KNOX* gene was obtained from the wheat genome database by TBtools software to analyze the cis-acting regulatory elements in the *KNOX* gene promoter. The obtained promoter region sequence was submitted to the PlantCARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) to identify the type and number of cis-acting elements.

3.6 Analysis of tissue expression pattern of *TaKNOX* gene family

Download transcriptome data in different periods of 5 different tissues including root, stem, leaf, spike and seed from the wheat gene expression website (http://www.wheat-expression.com/) (Ramirez-Gonzalez et al., 2018). The tissue expression data of *TaKNOX* gene were transformed and visualized by TBtools software.

Authors' contributions

ANN was the executor of the experimental design and experimental research in this study. SNN and LHH completed the data analysis and drafted the manuscript. AYW, BSL, MFF, ZZ, LH, ZY, and SCP participated in the design of the study and performed the analysis of the results. GGH conceived of the project, directed the design of the study, data analysis, draft and revision. All authors read and approved the final manuscript.

Acknowledgments

The study was supported by the National Natural Science Foundation of China (31901547), China Postdoctoral Foundation (2018M642741), and the Postdoctoral Research Project in Henan Province(001801013).

Reference

Cannon S.B., Mitra A., Baumgarten A., Young N.D., and May G., 2004, The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana*, BMC Plant Biol., 4: 10

Chen C.J., Chen H., Zhang Y., Thomas H.R., Frank M.H., He.Y., and Xia R., 2020, TBtools: An integrative toolkit developed for interactive analyses of big biological data, Mol. Plant, 13: 1194-202
https://doi.org/10.1016/j.molp.2020.06.009

Finn R.D., CoggillP., Eberhardt R.Y., Eddy S.R., Mistry J., Mitchell A.L., Potter S.C., Punta M., Qureshi M., Sangrador-Vegas A., Salazar G.A., Tate J., and Bateman A., 2016, The Pfam protein families database: towards a more sustainable future, Nucleic. Acids Res., 44: D279-285
https://doi.org/10.1093/nar/gkv1344

Gao J., Yang X., Zhao W., Lang T., and Samuelsson T., 2015, Evolution, diversification, and expression of KNOX proteins in plants, Front Plant Sci., 6: 882
https://doi.org/10.3389/fpls.2015.00882

Hamant O., and Pautot V., 2010, Plant development: a TALE story, C. R. Biol., 333: 371-381
https://doi.org/10.1016/j.crvi.2010.01.015

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

He Z.H., Zhang Q.S., Cheng S.H., Zhao Z.D., and Liu X., 2018, Wheat production and technology improvement in China, Nongxue Xuebao (Journal of Agriculture), 8(1): 99-106

Iwgsc, 2018, Shifting the limits in wheat research and breeding using a fully annotated reference genome, Science, 361: 661

Jia P., Zhang C., Xing L., Li Y., Shah K., Zuo X., Zhang D., An N., Han M., and Ren M., 2020, Genome-wide identification of the MdKNOX gene family and characterization of Its transcriptional regulation in malus domestica, Front Plant Sci., 11: 128
https://doi.org/10.3389/fpls.2020.00128

Krzywinski M., Schein J., Birol I, Connors J., Gascoyne R., Horsman D., Jones S.J., and Marr M.A., 2009, Circos: an information aesthetic for comparative genomics, Genome Res., 19: 1639-1645
https://doi.org/10.1101/gr.092759.109

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K., 2018, MEGA X: Molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol., 35: 1547-1549
https://doi.org/10.1093/molbev/msy096

Ma Q., Wang N.H., Hao P.B., Sun H.R., Wang C.C., Ma L., Wang H.T., Zhang X.L., Wei H.L., and Yu S.H., 2019, Genome-wide identification and characterization of TALE superfamily genes in cotton reveals their functions in regulating secondary cell wall biosynthesis, Bmc. Plant Biology, 19: 432
https://doi.org/10.1186/s12870-019-2026-1

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

Morimoto R., Kosugi T., Nakamura C., and Takumi S., 2005, Intragenic diversity and functional conservation of the three homoeologous loci of the KN1-type homeobox gene wknox1 in common wheat, Plant Mol. Biol., 57: 907-924
https://doi.org/10.1007/s11103-005-3247-2

Mukherjee K., Brocchieri L., and Burglin T.R., 2009, A comprehensive classification and evolutionary analysis of plant homeobox genes, Mol. Biol. Evol., 26: 2775-2794
https://doi.org/10.1093/molbev/msp201

Ramirez-Gonzalez R.H., Borrill P., Lang D., Harrington S.A., Brinton J., Venturini L., Davey M., Jacobs J., Fvan Ex., Pasha A., Khedikar Y., Robinson S.J., Cory A.T., Florio T., Concia L., Juery C., Steuernagel B., Xiang D., Ridout C.J., Chalhoub B., Mayer K.F.X., Benhamed M., Latrasse D., Bendahmane A., Consortium International Wheat Genome Sequencing, Wulff B.B.H., Appels R., Tiwari V., Datla R., Choulet F., Pozniak C.J., Provart N.J., Sharpe A.G., Paux E., Spannagl M., Brautigam A., and Uauy C., 2018, The transcriptional landscape of polyploid wheat, Science, 361: 1-3
https://doi.org/10.1126/science.aar6089

Su Y.H., Zhou C., Li Y.J., Yu Y., Tang L.P., Zhang W.J., Yao W.J., Huang R.F., Thomas L., and Zhang X.S., 2020, Integration of pluripotency pathways regulates stem cell maintenance in the Arabidopsis shoot meristem, P. Natl. A. Sci. India B., 117: 22561-22571
https://doi.org/10.1073/pnas.2015248117

Takumi S., Kosugi T., Murai K., Mori N., and Nakamura C., 2000, Molecular cloning of three homoeologous cDNAs encoding orthologs of the maize KNOTTED1 homeobox protein from young spikes of hexaploid wheat, Gene, 249: 171-181
https://doi.org/10.1016/S0378-1119(00)00164-5

Tsuda K., and Hake S., 2015, Diverse functions of KNOX transcription factors in the diploid body plan of plants, Curr. Opin. Plant Biol., 27: 91-96
https://doi.org/10.1016/j.pbi.2015.06.015

Tsuda K., Ito Y., Sato Y., and Kurata N., 2011, Positive autoregulation of a *KNOX* gene is essential for shoot apical meristem maintenance in Rice, Plant Cell, 23: 4368-4381
https://doi.org/10.1105/tpc.111.090050

Vollbrecht E., Veit B., Sinha N., and Hak S.E., 1991, 'The developmental gene Knotted-1 is a member of a maize homeobox gene family', Nature, 350: 241-243
https://doi.org/10.1038/350241a0

Wang S.M., Yamaguchi M.E., Grienenberger P.T., Martone A.L., Samuels, and Mansfield S.D., 2020, 'The Class II KNOX genes KNAT3 and KNAT7 work cooperatively to influence deposition of secondary cell walls that provide mechanical support to Arabidopsis stems', Plant Journal, 101: 293-309
https://doi.org/10.1111/tpj.14541

Wang Y.P., Tang H.D., Debarry J.D., Tan X., Li J.P., Wang X. Y., Lee T., Jin H.Z., Marler B., Guo H., Kissinger J.C., and Paterson A. H., 2012, MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity', Nucleic. Acids Res., 40: e49
https://doi.org/10.1093/nar/gkr1293

Xiong H.Z., Shi A., Wu D.X., Weng Y.J., Qin J, Ravelombola W.S., Shu X.L., and Zhou W., 2018, Genome-Wide identification, classification and evolutionary expansion of KNOX gene family in rice (*Oryza sativa*) and populus (Populustrichocarpa), American Journal of Plant Sciences, 9: 1071-1092
https://doi.org/10.4236/ajps.2018.96082

Xu S., Sun Y.M., Jia Z.C., and Luo K.M., 2009, Effects of overexpression of IPT and KN1 on development and growth in transgenic tobacco plants, Zhiwu Shenglixue Tongxun (Plant Physiology Communications), 45(6): 537-543