

分子植物育种
Molecular Plant Breeding
ISSN 1672-416X, CN 46-1068/S

《分子植物育种》网络首发论文

题目： 小麦 KNOX 基因家族鉴定与生物信息学分析
作者： 宋宁宁，梁慧慧，安义伟，柏胜龙，麻菲菲，张震，李浩，周云，郭光辉，
宋纯鹏
网络首发日期： 2021-01-05
引用格式： 宋宁宁，梁慧慧，安义伟，柏胜龙，麻菲菲，张震，李浩，周云，郭光辉，
宋纯鹏. 小麦 KNOX 基因家族鉴定与生物信息学分析. 分子植物育种.
<https://kns.cnki.net/kcms/detail/46.1068.S.20210105.1103.008.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

研究报告

Research Report

小麦 *KNOX* 基因家族鉴定与生物信息学分析

宋宁宁 梁慧慧 安义伟 柏胜龙 麻菲菲 张震 李浩 周云 郭光辉* 宋纯鹏

河南大学生命科学学院, 省部共建作物逆境适应与改良国家重点实验室, 开封, 475004

*通信作者, guanghuigu2008@163.com

摘要 *KNOX* 基因家族是编码同源异型盒蛋白的转录因子, 在植物生长发育和形态建成过程中起着重要的调控作用。然而, 关于小麦 *KNOX* 基因家族的研究很少。在本研究中, 利用生物信息学手段从小麦基因组中鉴定到分布于 18 条染色体上的 36 个 *KNOX* 基因, 所有鉴定到的家族成员均含有 *KNOX1* 或 *KNOX2* 结构域。我们对 *KNOX* 基因家族的系统进化、基因结构、蛋白结构域、顺式作用元件和基因表达模式进行了分析。根据其系统发育关系, 将小麦 *KNOX* 基因分为 Class I 和 Class II 两大亚类, 两亚类又进一步分为 5 个进化分支。大多数小麦 *KNOX* 基因含有典型 *KNOX1*、*KNOX2*、*ELK*、*HOX* 四个蛋白保守结构域。小麦 *KNOX* 基因的启动子分析发现了一些与激素、生长发育和胁迫响应相关的顺式作用调控元件。不同组织器官转录组数据分析显示, Class I 类 *KNOX* 基因具有明显的组织特异性, Class II 类 *KNOX* 基因的表达部位更广泛。本研究为进一步分析小麦 *KNOX* 基因家族的功能和调控机制提供了重要信息。

关键词 小麦; *KNOX* 基因家族; 基因结构; 基因表达

基金项目: 本研究由国家自然科学基金项目(31901547)、中国博士后基金(2018M642741)和河南省博士后科研项目(001801013)共同资助。引用格式: Song N.N., Liang H.H., An Y.W., Bai S.L., Ma F.F., Zhang Z., Li H., Zhou Y., Guo G.H., and Song C.P., 2020, Identification and bioinformatics analysis of *KNOX* gene family in wheat (*Triticum aestivum*L.), *ZhiwuFenziYuzhong* (Molecular Plant Breeding), (宋宁宁, 梁慧慧, 安义伟, 柏胜龙, 麻菲菲, 张震, 李浩, 周云, 郭光辉, 宋纯鹏, 2020, 小麦 *KNOX* 基因家族鉴定与生物信息学分析, 分子植物育种)

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

*Corresponding author, guanghui2008@163.com

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 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)转录因子家族是同源异型盒基因(Homeobox gene) TALE 蛋白超家族(Three amino acid loop extension superfamily)的一个亚族, 广泛参与调控植物的生长发育过程(Hamant and Pautot 2010)。大多数 *KNOX* 蛋白具有 4 个特征保守结构域, 分别是 KNOX1、KNOX2、ELK 和 HOX(Magnani and Hake 2008; Gao et al., 2015)。

KNOX 基因家族在植物界中是一个保守的基因家族, 从低等植物藻类、苔藓类到高等种子植物中都普遍存在(Gao et al. 2015)。植物中第一个被发现的 *KNOX* 基因是玉米的 Knotted1(*Kn1*)基因(Vollbrecht et al.

1991), 随后在越来越多的植物中鉴定出 *KNOX* 基因(Gao et al. 2015), 如拟南芥、水稻、杨树、棉花、苹果等, 并且部分 *KNOX* 基因功能已有较为深入的研究(Mukherjee et al. 2009; Xiong et al. 2018; Ma et al. 2019; Jia et al., 2020)。然而, 对重要粮食作物小麦 *KNOX* 基因报道较少, 目前只有 Takumi 等(2000)利用玉米的 *Kn1* 基因序列在小麦中同源克隆到了三个 *KNOX* 基因的报道。

根据 *KNOX* 基因的结构特征、系统发育树和表达模式, 可以将 *KNOX* 基因家族分为三类 Class I、Class II, 及双子叶植物中特有的 Class KNATM(Magnani and Hake 2008; Gao et al. 2015; Xiong et al. 2018)。不同类的 *KNOX* 基因的表达模式及生物学功能有着明显分化。Class I 类基因的表达较为集中, 主要在 SAM (Shoot apical meristem)中表达, 它们在分生组织的分化和维持中起着重要的作用(Tsuda et al. 2011; Gao et al. 2015; Su et al. 2020), 例如, 拟南芥 Class I 中 *STM* 基因对拟南芥胚胎中 SAM 的建立起着重要的作用, 且拟南芥 *stm* 突变体种子中由于缺少 SAM 只能产生子叶而不能产生新叶(Hay and Tsiantis 2010); Class I 中的 *KNOX* 基因与转基因过程中的愈伤分化有着密切的关系, 水稻 Class I 中 *osh1* 和 *osh15* 基因的双缺失突变体其愈伤组织只能形成叶状结构, 未见芽的形成(Tsuda et al. 2011); 将玉米 Class I 中的 *KN1* 基因转入烟草中, 其遗传转化成苗率与对照相比提高了一倍(胥珊等, 2009)。同时 Class I 类基因也在植物叶片形状的控制、节间伸长、激素平衡和花序结构的建立等方面发挥着重要的作用, 并可作为转录激活或抑制因子(Hay and Tsiantis 2010; Tsuda and Hake 2015); Class II 的 *KNOX* 亚家族基因表达部位一般比较广泛, 在植物的各个组织器官中都有所表达, 但是有关 Class II *KNOX* 基因在植物中的功能报道相对较少, 主要与次生细胞壁的形成相关。例如: *KNAT7* 与 *KNAT3* 协同作用影响次生细胞壁的沉积从而提高拟南芥茎的机械支持(Wang et al. 2020); Class KNATM 是双子叶植物特有的一类 *KNOX* 基因亚家族, 对拟南芥的 *KNATM* 基因研究发现该基因参与调节叶片极性和叶片形状(Magnani and Hake 2008)。

作为世界上最重要的粮食作物之一的小麦, 全球约 35%~40% 的人口以它作为主粮(何中虎等, 2018), 然而由于小麦基因组的复杂性, 对小麦 *KNOX* 基因克隆与研究的报道较少。随着小麦基因组信息的不断完善(Iwaguchi, 2018), 为在全基因组水平上鉴定、分析小麦 *KNOX* 基因带来契机。本研究利用小麦基因组信息, 结合 *KNOX* 基因家族特征, 对小麦 *KNOX* 基因进行了鉴定, 并对 *KNOX* 基因的基本理化性质、染色体分布、基因复制、进化关系、基因结构、顺式作用元件和不同组织的表达模式进行了分析, 为进一步研究小麦 *KNOX* 基因的功能和调控机制奠定了基础。

1 结果与分析

1.1 小麦 *KNOX* 基因的鉴定与命名

根据拟南芥和水稻 *KNOX* 蛋白的氨基酸序列在小麦数据库中进行 BLAST 搜索、筛选和验证，共鉴定到 36 个小麦 *KNOX* 基因。根据基因所在染色体位置命名为 *TaKNOX1*~*TaKNOX36*(表 1)。在 *TaKNOX* 基因家族编码的蛋白质中，氨基酸的数目从 153~389 个。*TaKNOX* 蛋白分子量从 16.54 到 42.48 kDa，等电点在 5.15~9.11，除 *TaKNOX13*、*TaKNOX15*、*TaKNOX18* 和 *TaKNOX34* 外，其余 *TaKNOX* 蛋白质的等电点都小于 7，说明大多数的 *TaKNOX* 蛋白质呈酸性。亚细胞定位预测分析表明，绝大多数 *KNOX* 基因定位在细胞核中，这与 *KNOX* 基因转录因子属性特征相符合。

表 1 小麦 *KNOX* 基因家族的理化性质

Table 1 Physicochemical properties of *TaKNOX* gene family

基因名	基因号	氨基酸数量(aa)	分子量(Da)	等电点	亚细胞定位
Gene name	Gene ID	Amino size(aa)	MW(Da)	PI	Subcellular localization
<i>TaKNOX1</i>	TraesCS1A02G072700	308	33778.98	5.66	细胞核; 细胞质-细胞核 Nucl;cyto_nucl
<i>TaKNOX2</i>	TraesCS1A02G072800	307	33144.26	5.40	细胞核 Nucl
<i>TaKNOX3</i>	TraesCS1B02G091600	308	33523.78	5.80	细胞核 Nucl
<i>TaKNOX4</i>	TraesCS1B02G091700	307	33202.34	5.40	细胞核 Nucl
<i>TaKNOX5</i>	TraesCS1B02G135600	282	31432.56	5.88	液泡; 叶绿素; 细胞核; 胞外; 高尔基体 vacu;chlo;nucl;extr;golg
<i>TaKNOX6</i>	TraesCS1D02G075600	307	33652.81	5.66	细胞核 Nucl
<i>TaKNOX7</i>	TraesCS1D02G075700	307	33174.29	5.40	细胞核 Nucl
<i>TaKNOX8</i>	TraesCS2A02G267400	348	38522.37	6.09	细胞核

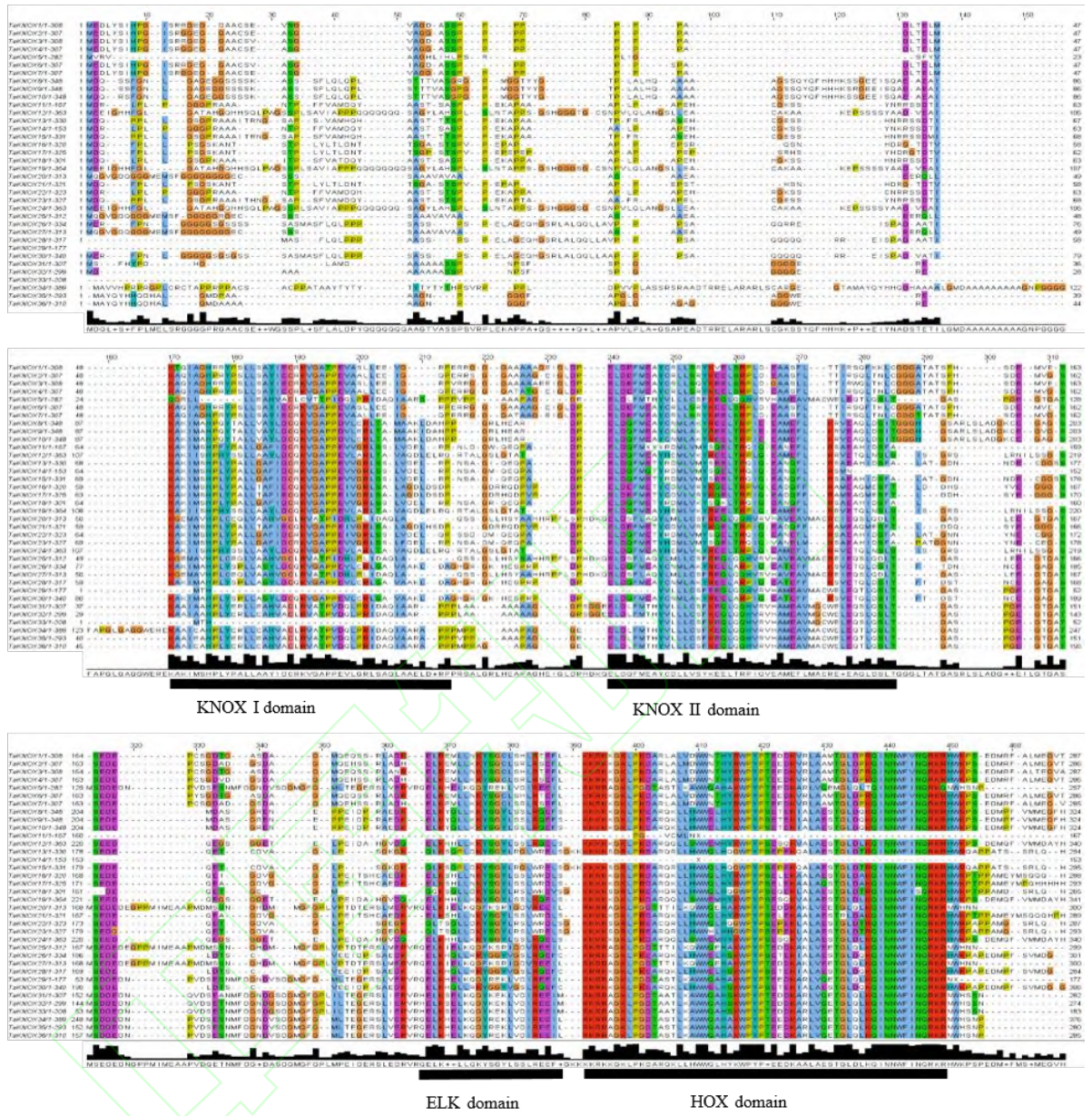
						Nucl
						细胞核
TaKNOX9	TraesCS2B02G268200	348	38560.47	6.09		Nucl
						细胞核
TaKNOX10	TraesCS2D02G256400	348	38463.30	5.98		Nucl
						线粒体; 叶绿素; 细胞核
TaKNOX11	TraesCS4A02G251800	167	18107.19	5.15		Mito;Chlo;Nucl
						细胞核
TaKNOX12	TraesCS4A02G256700	363	40079.08	6.28		Nucl
						细胞核
TaKNOX13	TraesCS4A02G291900	330	36155.39	8.61		Nucl
						叶绿素; 线粒体; 细胞核; 细胞核-质粒
TaKNOX14	TraesCS4A02G292000	153	16539.84	6.10		Chlo; Mito;nucl;Nucl_plas
						细胞核
TaKNOX15	TraesCS4A02G292100	331	36332.65	8.61		Nucl
						细胞核
TaKNOX16	TraesCS4A02G292200	320	35589.70	5.50		Nucl
						细胞核
TaKNOX17	TraesCS4B02G021900	325	36308.44	5.80		Nucl
						细胞核
TaKNOX18	TraesCS4B02G022000	301	33269.52	9.11		Nucl
						细胞核
TaKNOX19	TraesCS4B02G057900	364	40193.18	6.28		Nucl
						细胞核
TaKNOX20	TraesCS4B02G346100	313	34437.99	5.76		Nucl
						细胞核
TaKNOX21	TraesCS4D02G019600	341	38120.78	6.23		Nucl
						细胞核
TaKNOX22	TraesCS4D02G019800	323	35512.79	6.32		Nucl

TaKNOX23	TraesCS4D02G019900	327			细胞核
			35765.91	6.08	Nucl
TaKNOX24	TraesCS4D02G058000	363			细胞核
			40065.05	6.28	Nucl
TaKNOX25	TraesCS4D02G341000	312			细胞核
			34496.07	5.86	Nucl
TaKNOX26	TraesCS5A02G405900	334			细胞核; 细胞质-细胞核
			36772.57	5.97	Nucl;cyto_nucl
TaKNOX27	TraesCS5A02G515000	313			细胞核
			34437.99	5.76	Nucl
TaKNOX28	TraesCS5B02G410600	317			细胞核
			35314.09	5.96	Nucl
TaKNOX29	TraesCS5B02G454600	177			细胞核; 细胞质; 叶绿素; 质粒;
			20218.73	5.30	Nucl;Cyto;Chlo;Plas
TaKNOX30	TraesCS5D02G415900	340			细胞核
			37367.24	5.97	Nucl
TaKNOX31	TraesCS6A02G145500	307			细胞核
			33472.45	5.72	Nucl
TaKNOX32	TraesCS6B02G173900	299			细胞核
			32459.32	5.69	Nucl
TaKNOX33	TraesCS6D02G134800	208			细胞核; 细胞质-细胞核
			23540.29	6.10	Nucl;Cyto_nucl
TaKNOX34	TraesCS7A02G511800	389			细胞核
			42477.98	7.66	Nucl
TaKNOX35	TraesCS7B02G423100	293			细胞核
			32422.46	5.90	Nucl
TaKNOX36	TraesCS7D02G501000	310			细胞核
			33985.99	5.69	Nucl

1.2 小麦 KNOX 蛋白的多序列比对

多重序列比对发现小麦 KNOX 蛋白序列有 4 个相对保守的区域，即 KNOX1、KNOX2、ELK 和 HOX (图 1A)。KNOX1 和 KNOX2 结构域位于蛋白的 N 端，而 ELK 和 HOX 位于蛋白的 C 端。其中 HOX 结构域具有最高的保守性，表现出 TALE 同源异型盒蛋白超家族的典型结构特征，即在第 1 和第 2 个螺旋之间存在 3 个额外的氨基酸序列(P-Y-P)。另外的 KNOX1、KNOX2 和 ELK 保守结构域也表现出较好的保守性。例如，保守结构域 KNOX2 具有高保守性的 E-L-D 氨基酸序列，保守结构域 ELK 具有高保守性的 E-L-K 氨基酸序列(图 1B)。

A



B

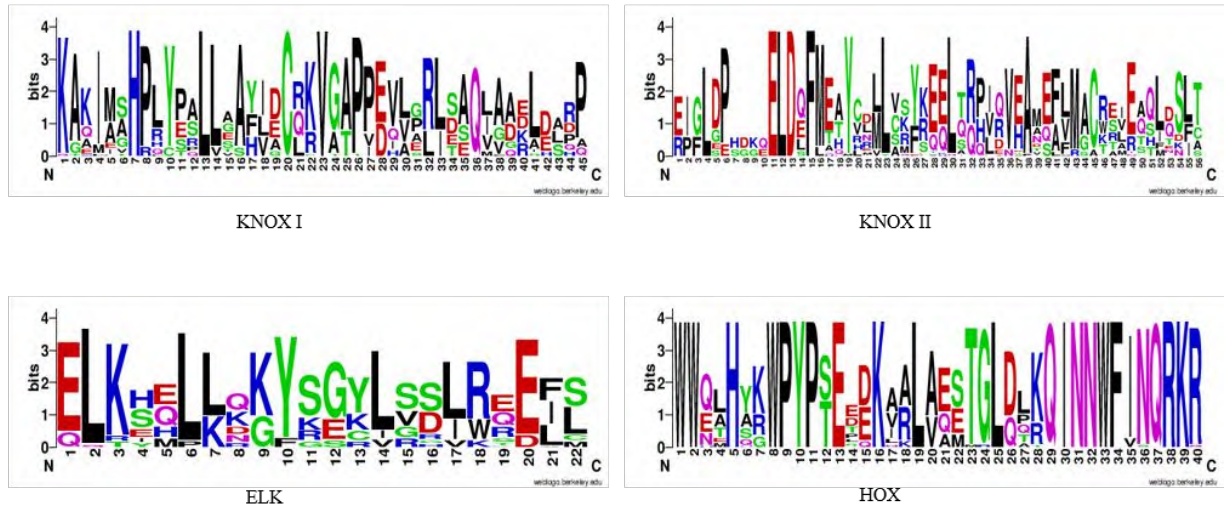


图 1 小麦 KNOX 蛋白多序列比对

注: A 图中黑色横线表示 KNOX I, KNOX II, ELK 和 HOX 四个保守结构域; B 图是每个保守结构域的蛋白特征基序的分布和组成; 字母符号代表氨基酸; 字母高度代表保守水平

Figure 1 Multiple alignment 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

1.3 小麦 *KNOX* 基因的染色体定位和基因复制分析

如图 2 所示, *KNOX* 基因除了在小麦的 3 号染色体无分布外, 其他染色体均有分布, 其中 2、6、7 号染色体上各有 1 个 *KNOX* 基因; 4A 染色体上 *KNOX* 基因最多, 含有 6 个 *KNOX* 基因; 同时, 大多数 *KNOX* 基因分布在染色体的端部。片段复制和串联重复是基因家族扩张的主要机制(Cannon et al. 2004), 对 *TaKNOX* 基因进行共线性分析发现, 其含有 5 对串联重复基因, 分别是 *TaKNOX1* 和 *TaKNOX2*、*TaKNOX3* 和 *TaKNOX4*、*TaKNOX6* 和 *TaKNOX7*、*TaKNOX14* 和 *TaKNOX15*、*TaKNOX22* 和 *TaKNOX23* 为串联重复基因(图 2 红色字体为串联重复基因); 另外发现有 23 对 *TaKNOX* 基因位于片段性共线性区块内, 其中包括大多数具有部分同源关系的 *TaKNOX* 基因(图 2 蓝色连线的为片段重复区块内的 *KNOX* 基因), 表明小麦 *KNOX* 基因的扩张存在串联重复和和多倍化过程中片段重复现象。

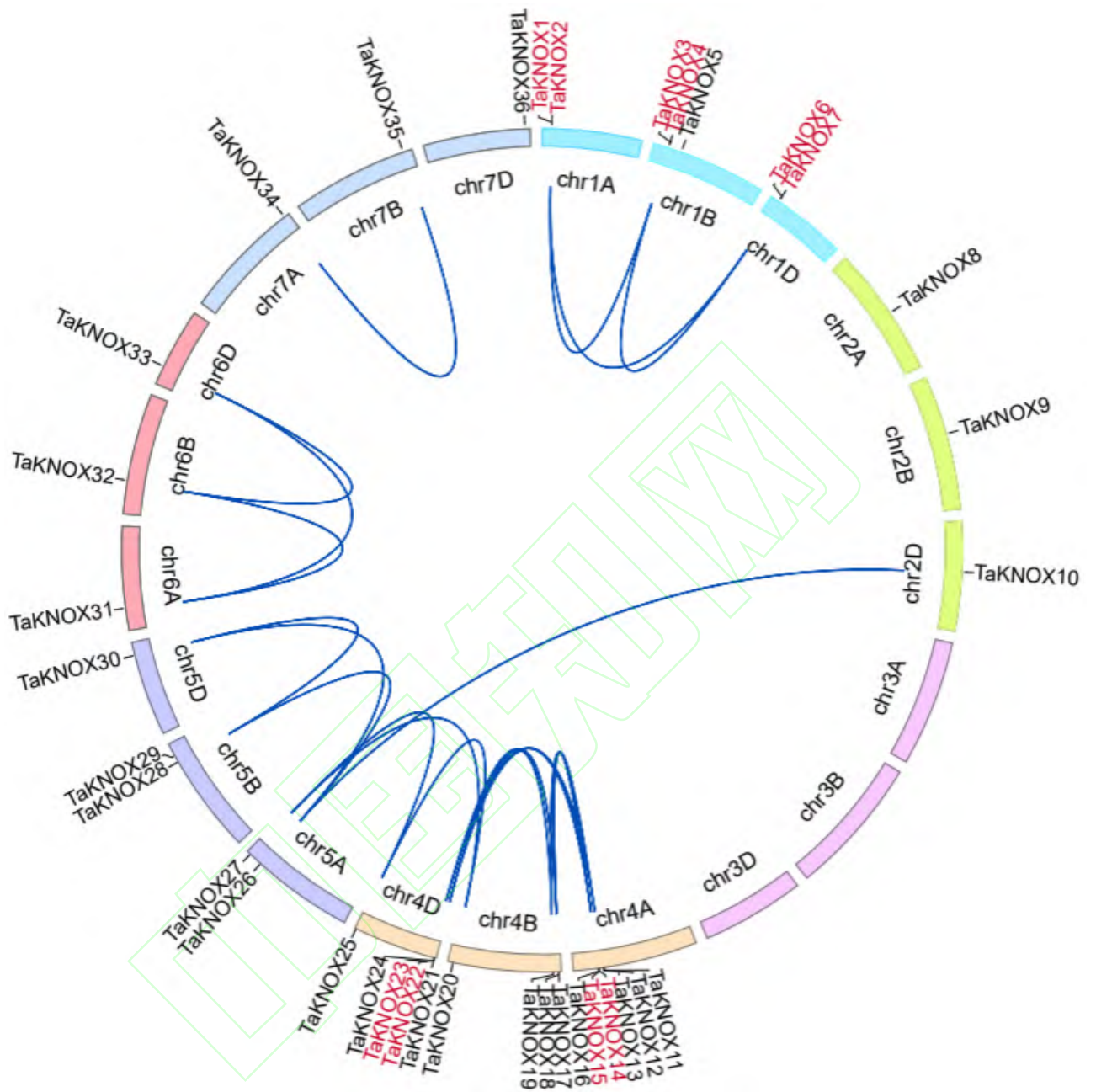


图 2 小麦 *KNOX* 基因染色体定位和基因复制

注: 蓝色连线代表在片段性重复区块内的小麦 *KNOX* 基因对; 红色字体为串联重复的小麦 *KNOX* 基因

Figure 2 Chromosome mapping and gene duplication of *KNOX* genes in 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 小麦 *KNOX* 基因的系统发育分析

将已知模式植物拟南芥和水稻的 *KNOX* 蛋白序列和小麦的 *KNOX* 蛋白序列, 使用 MEGA X 软件构建

KNOX 系统发育树(图 3)。根据系统发育树, *KNOX* 基因被分为三大类: Class I、Class II 和 Class KNATM。其中, I 类基因 38 个, 包括 4 个拟南芥基因, 9 个水稻基因和 25 个小麦基因; II 类基因 19 个, 其中拟南芥基因 4 个, 水稻基因 4 个, 小麦基因 11 个; Class KNATM 类基因只包含了拟南芥的一个成员 *KNATM*, 该类 *KNOX* 基因是双子叶植物特有的一类(Magnani and Hake, 2008)。

Class I 又进一步分为三个亚类, 分别是 Class I A、Class I B 和 Class I C。Class II 包括 Class II A 和 Class II B 两个亚类(图 3)。Class I A 中 *TaKNOX19*、*TaKNOX24* 和 *TaKNOX26* 与拟南芥的 *AtKNAT1* 和 *AtSTM*、水稻的 *LOC_Os03g51690(OSH1)* 聚为一支; Class I B 中 *TaKNOX1*、*TaKNOX2*、*TaKNOX3*、*TaKNOX4*、*TaKNOX6* 和 *TaKNOX7* 与拟南芥的 *AtKNAT2* 和 *AtKNAT6* 聚到一起。Class II A 中的 *TaKNOX20*、*TaKNOX25* 和 *TaKNOX27* 同拟南芥的 *AtKNAT7* 聚为一支(图 3), 这些小麦 *KNOX* 基因与其系统发育关系较近的拟南芥、水稻 *KNOX* 基因可能具有相似的生物学功能。同时, 位于 A、B、D 亚基因组上的 *KNOX* 部分同源基因亲缘关系最近, 它们大多数聚在一起, 如 *TaKNOX1*、*TaKNOX3*、*TaKNOX6*、*TaKNOX2*、*TaKNOX4*、*TaKNOX7*、*TaKNOX8*、*TaKNOX9*、*TaKNOX10*、*TaKNOX12*、*TaKNOX19*、*TaKNOX24*、*TaKNOX26*、*TaKNOX28*、*TaKNOX30* 等。

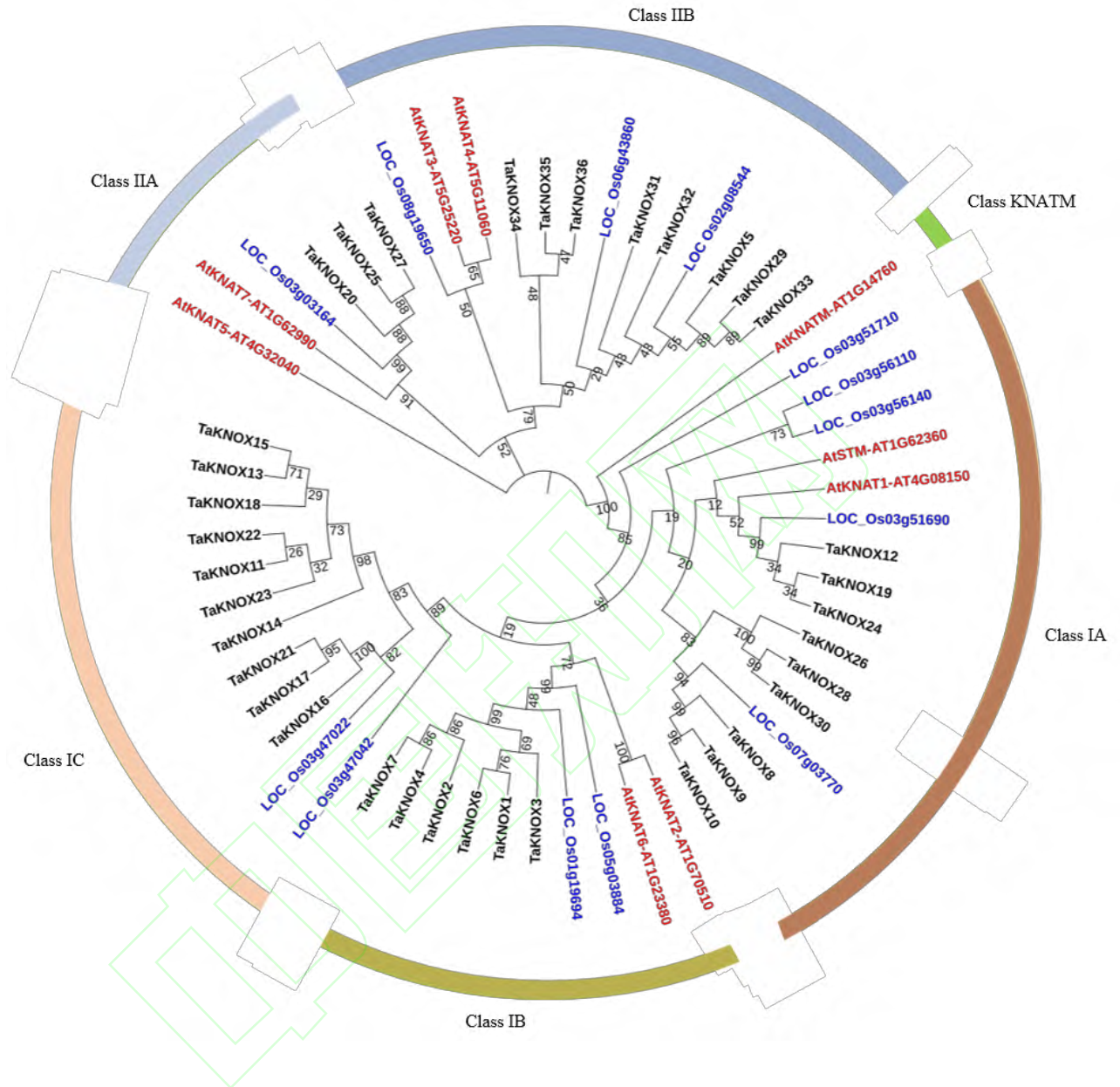


图 3 小麦、拟南芥和水稻 *KNOX* 基因家族的系统进化树

注: 黑色, 红色和蓝色: 小麦、拟南芥和水稻 *KNOX* 基因

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.

1.5 小麦 *KNOX* 基因结构和蛋白保守结构域分析

为了进一步探究小麦 *KNOX* 基因家族的特征, 我们基于种内系统发育树(图 4A)分析了 *TaKNOX* 基因的基因结构和蛋白保守结构域分布(图 4)。结果显示, *TaKNOX* 家族主要有 3~6 个外显子和 4~5 个内含子组成。Class I 中的基因外显子数量在 4~6 之间, 而 Class II 中的基因几乎都是含有 5 个外显子(图 4B)。值

值得注意的是，Class I 类基因中除 TaKNOX11 和 TaKNOX14 外，都含有一个较长的内含子(图 4B)，这与 Class I 类 *KNOX* 基因家族基因结构特征相符合(Morimoto et al., 2005)。大多数 TaKNOX 蛋白包含 KNOX1、KNOX2、ELK 和 HOX 保守的蛋白结构域，并在蛋白序列上依次排列；个别 TaKNOX 蛋白缺少 1~2 个保守结构域，如 TaKNOX11、TaKNOX14 缺少 ELK 和 HOX 保守域，TaKNOX29、TaKNOX33 缺少 KNOX1 结构域，但所有的 TaKNOX 蛋白都含有 KNOX2 保守结构域(图 4C)。同时，从进化关系上看，亲缘关系较近的同亚家族的 *TaKNOX* 基因在基因结构组成和蛋白结构域分布上也是保守的。

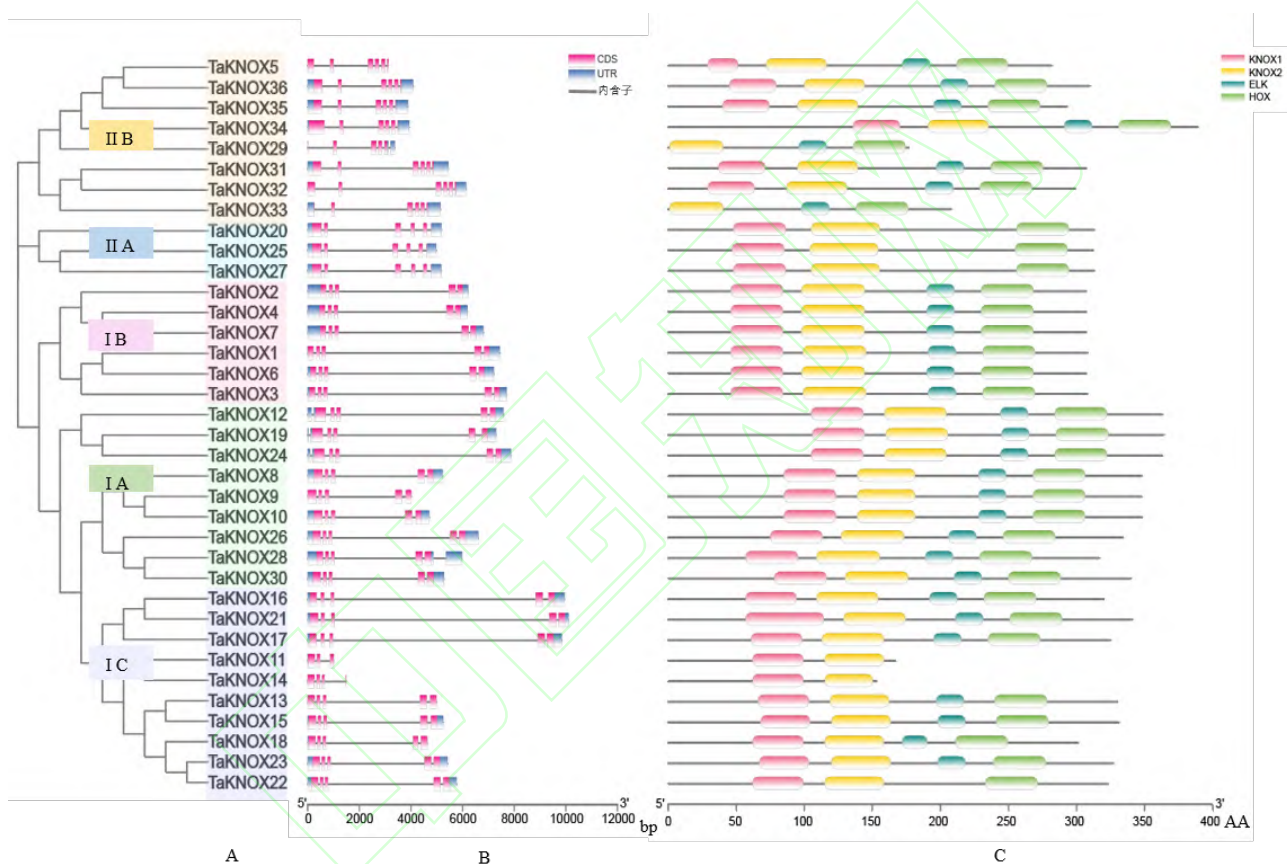


图 4 小麦 *KNOX* 基因家族的基因结构和蛋白保守结构域分析。

注: A: 小麦 *KNOX* 基因家族的系统发育树; B: 小麦 *KONX* 基因家族的基因结构组成; C: 小麦 *KNOX* 基因家族的保守结构域组成

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 小麦 *KNOX* 基因启动子上顺式作用元件分析

为了分析小麦 *KNOX* 基因家族成员潜在的表达调控机制，我们对每个成员的顺式作用元件进行鉴定。

对鉴定的顺式作用元件，选择与植物激素、植物生长、参与胁迫反应和光响应相关的作用元件进行分析(图

5)。结果表明，所有 *TaKNOX* 基因的启动子区均含有与植物激素和胁迫响应相关的元件，例如有关植物激素相关的顺式作用元件有 ABA 响应元件 ABRE，MeJA 响应元件 CGTCA-motif、TGACG-motif，GA 响应元件 TATC-box、GARE-motif、P-box 等；有关胁迫响应相关的顺式作用元件有低温诱导响应元件 LTR，防御和应激反应相关的响应元件 TC-rich repeats 等；另外，28 个 *TaKNOX* 基因都含有与植物生长发育调控相关的顺式元件(O₂-site, CAT-box, GCN4-motif, circadian),值得注意的是，有 23 个 *TaKNOX* 基因含有与植物分生组织表达相关的元件(CAT-box)，这些结果表明 *TaKNOX* 基因可能在小麦的生长发育、分生组织分化能力的维持和响应逆境胁迫方面起着重要的作用(图 5)。

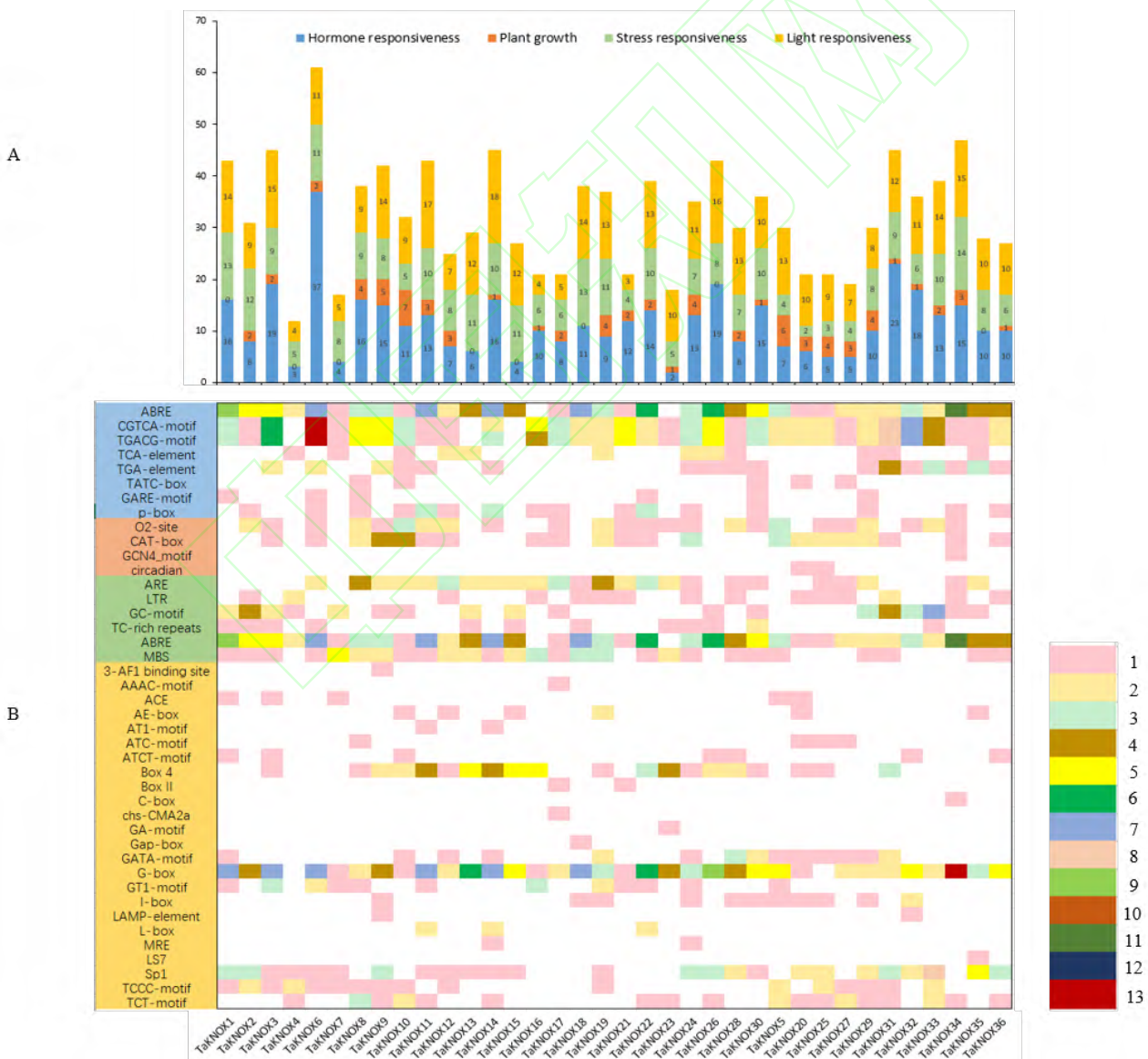


图 5 小麦 *KNOX* 基因顺式作用元件分析。

注: A: 用不同颜色的直方图表示 *TaKNOX* 基因中四大类顺式作用元件的总和; B: 每个 *TaKNOX* 基因顺式作用元件数量的热图; 不同的颜色表示顺式作用元件的数量; 白色代表没有顺式作用元件

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* genes. Different colors indicate the number of cis-acting elements, while white indicates that there are no cis-acting elements

1.7 小麦 *KNOX* 基因在不同组织中的表达模式

为了探讨不同生长发育阶段 *KNOX* 基因在小麦中的组织特异性表达模式，对 36 个小麦 *KNOX* 基因的表达数据进行了分析，研究了小麦根、茎、叶、穗和种子等不同组织的转录水平，发现小麦 *KNOX* 基因在不同的组织或器官中的表达模式有明显的差异(图 6)。其中小麦 *KNOX* 基因的表达模式与其亚家族分类呈现出一定的关联性。Class I 中，除 *TaKNOX18* 主要在叶片中表达，其他的成员主要在茎和穗中表达，且 *TaKNOX11*、*TaKNOX13*、*TaKNOX14* 和 *TaKNOX15* 特异性在穗中高表达；而 Class II 中，除 *TaKNOX34*、*TaKNOX35*、*TaKNOX36* 在叶片中特异高表达外，其他基因表达部位则比较广泛。

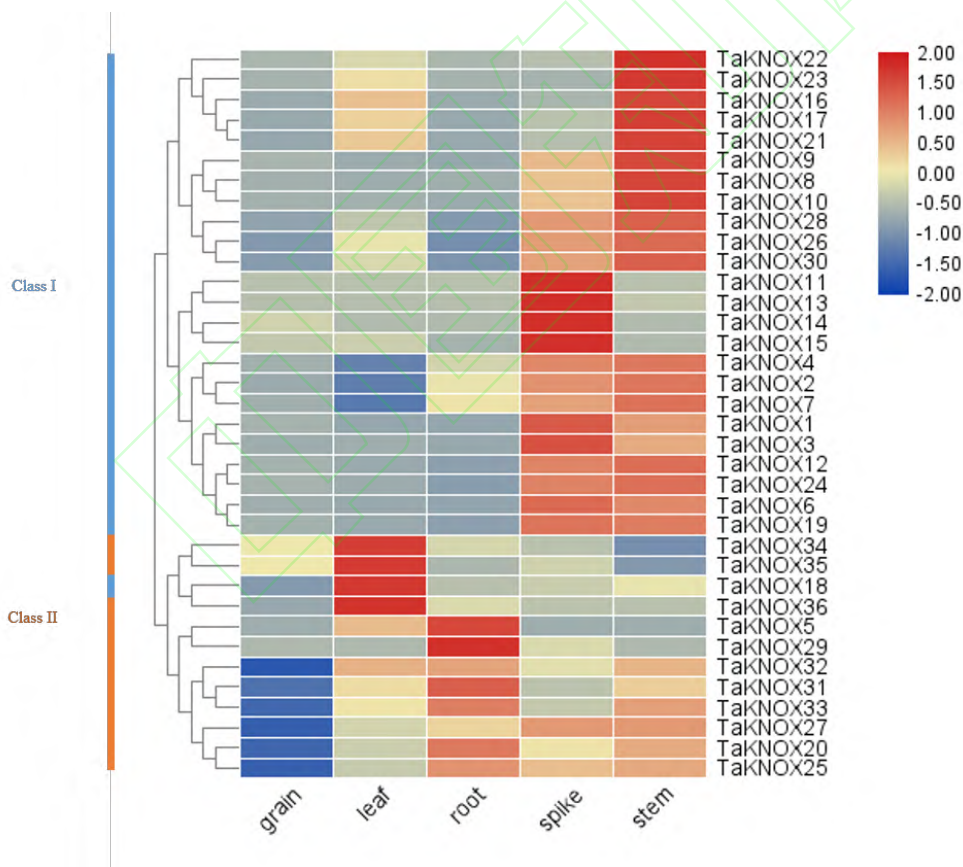


图 6 小麦 *KNOX* 基因组织表达热图

Figure 6 *TaKNOX* gene expression profiles in different tissues.

2 讨论

KNOX 基因作为重要的植物生长发育调控因子，在小麦中的鉴定与功能研究十分欠缺。目前有关小麦

KNOX 基因的研究报道较少, Takumi 等利用玉米的 *Kn1* 基因序列在小麦中同源克隆到了三个 *KNOX* 基因 (Takumi et al. 2000), 即为本研究中的 *TaKNOX12*、*TaKNOX19* 和 *TaKNOX24*, Morimoto 等(2005)对这三个小麦 *KNOX* 基因的表达模式和转基因烟草表型进行了分析。随着小麦基因组信息的不断完善, 这为从全基因组水平鉴定分析 *TaKNOX* 转录因子带来契机。本研究首次在全基因组水平上对小麦的 *KNOX* 基因家族进行了鉴定, 并对小麦 *KNOX* 基因的基本理化性质, 基因结构, 进化关系, 表达模式等进行了分析, 为后续的 *KNOX* 基因的功能研究奠定了基础。

本研究中, 我们共鉴定到了 36 个 *KNOX* 基因家族成员, 数目多于拟南芥(9 个)和水稻(13 个)。在这些基因中, 有 5 对串联重复和 28 对片段复制, 说明全基因组重复时间在小麦 *KNOX* 基因家族扩增过程中发挥了重要作用, 因此导致小麦 *KNOX* 基因家族比拟南芥和水稻庞大。根据蛋白序列、进化树和基因结构特征把小麦 *KNOX* 基因分成了两大类 Class I 和 Class II。其中 *TaKNOX1~4*、*TaKNOX6~19*、*TaKNOX21~24*、*TaKNOX26*、*TaKNOX28* 和 *TaKNOX30* 与拟南芥和水稻中 Class I 类基因存在于同一进化支, 剩余的 *TaKNOX* 基因与拟南芥和水稻中的 Class II 类基因聚为一支。值得注意的是 Class I 中的 *TaKNOX12*、*TaKNOX19*、*TaKNOX24* 与拟南芥 *AtSTM* (*AT1G6230*)、*AtKNAT1* (*AT4G08150*), 及水稻的 *OSHI* (*LOC_Os03g51690*) 亲缘关系最近, 这三个 *KNOX* 基因可能与 *AtSTM*、*AtKNAT1* 和 *OSHI* 具有相似的生物学功能, 即影响胚胎中 SAM 的建立、胚性愈伤组织的形成(Hay and Tsiantis 2010; Tsuda et al. 2011), 因此小麦这三个 *KNOX* 基因或可应用于提高小麦遗传转化过程中胚性愈伤组织形成的效率, 从而提高小麦的遗传转化效率。

小麦 *KNOX* 基因转录组数据结果显示, 除 *TaKNOX18* 基因, Class I 类基因的表达均比较集中, 除了在穗和茎中表达水平很高之外, 在其他组织中的表达水平均很低或者不表达; 而 Class II 类基因几乎在小麦的所有组织中均有表达, 这与其他植物 *KNOX I* 和 *KNOX II* 类基因的表达模式相符合, 表明植物 *KNOX* 蛋白在功能上可能具有高度的保守性(Gao et al. 2015)。值得一提的是, Class I 类基因中的 *TaKNOX8*、*TaKNOX9*、*TaKNOX10*、*TaKNOX12*、*TaKNOX19* 和 *TaKNOX24* 在小麦茎中的表达量是最高的, 且在茎发育早期含量最高, 随着发育后期含量逐渐下降, 这可能说明这些基因在小麦茎端分生组织的维持和调控中起着重要的作用, 关于这些基因的功能有待于进一步研究。因此深入了解小麦 *KNOX* 基因的功能, 将为研究小麦的生长发育、形态建成等过程奠定重要基础。

3 材料与方法

3.1 小麦 *KNOX* 基因家族的鉴定

鉴定小麦基因组数据库中所有 *KNOX* 基因，主要步骤如下：(1)从 Ensemble Plants 数据库 (http://plants.ensembl.org/Triticum_aestivum/Info/Index) 下载小麦中国春的基因组数据；(2)分别从 TAIR (<https://www.arabidopsis.org/>) 网站和 Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/index.shtml>) 网站下载拟南芥的 9 条 *KNOX* 蛋白序列和水稻的 13 个 *KNOX* 蛋白序列；(3)以拟南芥和水稻的 *KNOX* 蛋白序列作为种子序列，利用 Tbttools(Chen et al. 2020)的 blast 命令在小麦蛋白数据库中进行搜索，设置阈值为 $e\text{-value} < e^{-10}$ ；(4)将搜索到的结果提交到“PFAM”网站 (<http://pfam.xfam.org/>)对 *KNOX* 蛋白保守结构域 *KNOX* I 或 *KNOX* II 的 Pfam 号(PF03790 或 PF03791)进行进一步的验证(Finn et. al. 2016)，含有 *KNOX* I 或 *KNOX* II 结构域的被视为 *KNOX* 基因家族成员。利用 ProtParam(<https://web.expasy.org/protparam/>)网站预测每个 *KNOX* 蛋白的氨基酸数(AA)、分子量(MW)和等电点(PI),并在 WoLF PSORT(<https://wolfsort.hgc.jp/>)网站上对 *KNOX* 蛋白的亚细胞定位进行预测。

3.2 小麦 *KNOX* 基因家族成员的蛋白结构域分析和多序列比对

采用 CDD 网站(<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)和 Pfam 网站(<http://pfam.xfam.org/>)对小麦的 *KNOX* 蛋白结构域进行分析。使用 MEGA X(Kumar et al. 2018)中的 MUSCLE 对 *KNOX* 蛋白序列进行多重序列比对，利用 Jalview 软件进行可视化。

3.3 小麦 *KNOX* 基因家族成员的染色体定位与基因复制

根据小麦基因组注释文件获取 *KNOX* 基因家族成员的染色体位置信息；利用 MCScanX 软件(Wang et al. 2012)计算并获取共线性区块及基因的串联重复信息；利用 Circos 软件(Krzywinski et al. 2009)对小麦 *KNOX* 基因的染色体位置信息、共线性关系、串联重复信息进行可视化。

3.4 *KNOX* 基因家族的系统发育分析

将拟南芥、水稻和小麦的 *KNOX* 基因家族成员利用 MEGA X 软件构建系统发育树。运用 MUSCLE 的比对方法对 *KNOX* 蛋白序列进行多重比对，选择邻接法(Neighbor-joining)构建系统发育树，具体参数如下：

采用自展法(Bootstrap method), 参数为 1000, 模型选择泊松模型(Poisson model)。

3.5 小麦 *KNOX* 基因家族成员的启动子区顺式作用元件分析

利用 TBtools 软件从小麦基因组数据库中获得每个 *KNOX* 基因的 CDS 序列上游 2000bp 的启动子区域, 来分析 *KNOX* 基因启动子中的顺式作用调控元件。将获得的启动子区域序列提交到 PlantCARE 数据库中 (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)来鉴定顺式作用元件的类型和数量。

3.6 小麦 *KNOX* 基因家族的组织表达模式分析

在小麦基因表达网站(<http://www.wheat-expression.com/>)下载根、茎、叶、穗和种子五种不同组织的不同时期转录组数据(Ramirez-Gonzalez et al. 2018)。对小麦 *KNOX* 基因的组织表达数据用 TBtools 软件进行对数转换和可视化。

作者贡献

宋宁宁是本研究的实验设计和实验研究的执行人; 宋宁宁和梁慧慧完成数据分析, 论文初稿的写作; 安义伟、柏胜龙、麻菲菲、张震、李浩、周云及宋纯鹏参与实验设计, 实验结果分析; 郭光辉是项目的构思者及负责人, 指导实验设计, 数据分析, 论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由国家自然科学基金项目(31901547)、中国博士后基金(2018M642741)和河南省博士后科研项目(001801013)共同资助。

参考文献

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.

Finn R.D., Cogill P., 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.
- 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.
- Hamant O., and Pautot V., 2010, Plant development: a TALE story, *C. R. Biol.*, 333: 371-381.
- Hay A., and Tsiantis M., 2010, KNOX genes: versatile regulators of plant development and diversity, *Development*, 137: 3153-3165.
- He Z.H., Zhang Q.S., Cheng S.H., Zhao Z.D., and Liu X., 2018, Wheat production and technology improvement in China, *NongxueXuebao (Journal of Agriculture)*, 8(1): 99-106. (何中虎, 庄巧生, 程顺和, 于振文, 赵振东, 刘旭, 2018, 中国小麦产业
发展与科技进步, *农学学报*, 8(1): 99-106.)
- Iwgc, 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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., Provarnt 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.
- 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.
- 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.
- 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.
- 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.
- Vollbrecht E., Veit B., Sinha N., and Hake S.E., 1991, 'The developmental gene Knotted-1 is a member of a maize homeobox gene family', *Nature*, 350: 241-243.
- Wang S.M., Yamaguchi M.E., Grienberger 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.
- 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.
- 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 (*Populus trichocarpa*), *American Journal of Plant Sciences*, 9: 1071-1092.
- 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. (胥珊, 孙一铭, 贾之春, 罗克明, 2009, 超量表达IPT和KN1基因对转基因烟草生长发育的影响, *植物生理学通讯*, 45(6): 537-543.)