

萝卜 TALE 转录因子家族的鉴定与分析

赵伟, 李锡香, 王海平, 贾会霞, 宋江萍, 阳文龙, 张晓辉

中国农业科学院蔬菜花卉研究所 农业农村部园艺作物生物学与种质创制综合重点实验室 农业农村部蔬菜作物基因资源与种质创制北京科学观测实验站, 北京 100081

赵伟, 李锡香, 王海平, 贾会霞, 宋江萍, 阳文龙, 张晓辉. 萝卜 TALE 转录因子家族的鉴定与分析. 生物工程学报, 2022, 38(1): 343-358.

ZHAO W, LI XX, WANG HP, JIA HX, SONG JP, YANG WL, ZHANG XH. Identification and analysis of the TALE transcription factor family in radish. Chin J Biotech, 2022, 38(1): 343-358.

摘要: TALE (three-amino acid loop extension) 转录因子在植物生长发育及细胞分化过程中起重要作用。在多种植物中均已鉴定出 TALE 转录因子的家族成员, 但是萝卜 TALE 转录因子家族的研究鲜有报道。文中通过生物信息学手段在象牙白萝卜全基因组中鉴定出了分布于 9 条染色体上的 33 个 TALE 家族基因。研究结果显示, 该家族中除与拟南芥 *KNATM* 同源的基因 *Rsa10037940* 外, 其余基因均含有编码 HOX 保守结构域的序列。这些基因含有 4-6 个外显子。萝卜的 33 个 TALE 基因与拟南芥中的 17 个同源基因存在共线性关系。33 个 TALE 基因启动子区的顺式元件中包含大量逆境响应元件。表达特性分析显示, 该基因家族 *BELL* 亚家族内有 4 个基因在根内表达量较高, *KNOX* 亚家族内有 2 个基因在茎和愈伤中表达量较高。该家族不同亚型成员之间的蛋白 3D 结构高度相似。编码蛋白均为弱酸、有亲水性。萝卜 TALE 基因家族在进化上较为保守, 分化上与拟南芥存在一致性, 与水稻差异较大。本研究为开展萝卜中 TALE 转录因子的生物学功能研究提供了重要参考。

关键词: 萝卜; TALE 基因家族; 基因结构; 进化分析; 组织表达

Received: April 29, 2021; **Accepted:** September 17, 2021; **Published online:** October 5, 2021

Supported by: National Natural Science Foundation of China (31772301); National Key Research and Development Program of China (2016YFD0100204-07); Fundamental Research Funds (IVF-BRF2020001)

Corresponding author: ZHANG Xiaohui. Tel: +86-10-82105947; E-mail: zhangxiaohui01@caas.cn

基金项目: 国家自然科学基金 (31772301); 国家重点研发计划 (2016YFD0100204-07); 基本科研业务费 (IVF-BRF2020001)

Identification and analysis of the TALE transcription factor family in radish

ZHAO Wei, LI Xixiang, WANG Haiping, JIA Huixia, SONG Jiangping, YANG Wenlong, ZHANG Xiaohui

Beijing Research Station of Vegetable Crop Gene Resource and Germplasm Enhancement, Ministry of Agriculture and Rural Affairs of People's Republic of China, Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture and Rural Affairs of People's Republic of China, Institute of Vegetables and Flowers, Chinese Academy of Agriculture Sciences, Beijing 100081, China

Abstract: Three-amino acid loop extension (TALE) transcription factors play important roles in plant growth and cell differentiation. There are plenty of studies on *TALE* transcription factors in several model plants, but not in radish (*Raphanus sativas*). A genome-wide bioinformatics analysis identified 33 *TALE* family genes in the Xiang-Ya-Bai (XYB) radish. These genes are distributed on nine chromosomes and all contain 4-6 exons. The 33 *TALE* genes in radish showed a co-linearity relationship with the 17 homologous genes in *Arabidopsis thaliana*. Moreover, a large number of stress response cis-elements were found in the promoter regions of these genes. Expression analysis showed that four genes in the *BELL* subfamily were highly expressed in roots, and two genes in the *KNOX* subfamily were highly expressed in shoots of bolting plants and callus. All radish *TALE* genes contain sequences encoding the conserved HOX domain, except for the gene *RSA10037940*, which is homologous to *Arabidopsis KNATM*. The deduced 3D structures of the TALE proteins irrespective of subtypes are highly similar. All the encoded proteins were weakly acidic and hydrophilic. The radish *TALE* gene family is relatively evolutionarily conserved, which was consistent with results from *Arabidopsis*, but quite different from that of rice. This study provides important clues for studying the biological functions of TALE transcription factors in radish.

Keywords: *Raphanus sativus*; *TALE* gene family; gene structure; evolutionary analysis; tissue expression

同源异型盒 (*Homeobox*) 基因是指能够编码产生保守同源异型盒结构域的转录因子编码基因。一个典型的同源异型结构域由 60 个氨基酸组成, 60 个氨基酸形成了 3 个螺旋区域, 第一和第二螺旋形成环状结构, 第二和第三螺旋形成螺旋-角-螺旋的结构^[1]。*Homeobox* 基因最早发现于果蝇的第 3 条染色体上, 随后证实该类基因广泛存在于动物和植物中, 果蝇的 *Antp* 基因和玉米的 *Knotted-1* 基因分别为动物和植物中最早被克隆的基因。在植物中, 根据结构

特征可将 *Homeobox* 转录因子家族成员分为 14 类: HD-ZIP I、HD-ZIP II、HD-ZIP III、HD-ZIP IV、PLINC、WOX、DDT、PHD、NDX、LD、PINTOX、SAWADEE、KNOX 和 BELL; 在 PlantTFDB 中, *Homeobox* 转录因子家族被分为 5 个家族: HD-ZIP、TALE、WOX、HB-PHD 和 HB-other^[2]。其中, TALE 转录因子家族包含 KNOX 亚家族和 BELL 亚家族, 编码 1 个由 63 个氨基酸组成的非典型同源结构域, 额外的 3 个氨基酸 (P, Y, P) 残基连接在第一螺

旋区域和第二螺旋区域之间。KNOX 蛋白基本包含 KNOX1、KNOX2、ELK 和 Homeodomain 4 个结构域,拟南芥中 KNATM 蛋白不含 Homeodomain 结构域但属于 KNOX 家族。研究表明,含 *KNOX1* 结构域的基因在分生组织中表达,对于分生组织的发育起着重要作用;含 *KNOX2* 结构域的基因在调控植物细胞壁的次生长以及根、茎、种皮、心材的发育方面发挥重要作用^[3]。BELL 蛋白包含 POX 和 Homeodomain 2 个结构域,在胚珠发育、叶柄发育和果实形成过程中起着重要作用^[4]。

随着生物信息学和植物全基因组测序的快速发展,杨树^[5]、辣椒^[6]、百脉根^[7]、亚洲棉^[8]等物种基因组中分别鉴定到 35、20、40、46 个 TALE 转录因子家族成员。杨树 TALE 转录因子家族成员中发现了 11 个盐胁迫响应基因,其中 *ptTALE5* 经过胁迫处理后上调表达,可能在应对盐胁迫中具有重要功能。辣椒 TALE 转录因子家族各亚族之间保守基序几乎完全一致,表明辣椒 TALE 转录因子家族进化过程高度保守。百脉根中鉴定出的 TALE 转录因子家族成员存在多条结构域缺失的序列,表明百脉根在进化过程中出现假基因化现象,可能某些基因会出现功能缺失。亚洲棉 TALE 转录因子家族的研究中发现部分 TALE 成员在棉纤维次生细胞壁(secondary cell wall, SCW)增厚期表达,为进一步了解纤维 SCW 生物合成网络提供了线索。

萝卜(*Raphanus sativas*)属于十字花科萝卜属,是重要的蔬菜作物。研究萝卜 TALE 转录因子的功能和调控机制有助于萝卜的功能基因组研究和分子育种。本研究以象牙白萝卜为实验材料,在全基因组范围内识别和鉴定了 TALE 基因家族成员,并对其基因结构、在基因组中的位置、与近缘植物中同源基因的共线性关系、启动子区顺式作用元件、组织表达特征、编码

蛋白的理化性质和模拟蛋白质结构、系统发育关系等进行分析,研究结果将为萝卜 TALE 转录因子的生物学功能解析提供有用信息。

1 材料与方法

1.1 数据来源

通过植物转录因子数据库 (PlantTFDB, <http://planttfdb.cbi.pku.edu.cn/>) 下载萝卜 (*Raphanus sativus*)、拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa* subsp. *japonica*) 和胡杨 (*Populus euphratica*) 的 TALE 转录因子信息,象牙白萝卜全基因组测序由中国农业科学院蔬菜花卉研究所等单位协作完成^[9]。

1.2 萝卜 TALE 基因家族的鉴定

本研究从 PlantTFDB4.0 中下载得到萝卜 TALE 转录因子蛋白序列,利用在线工具 ClustalW 进行多序列比对获得 ALN 格式文件,通过 Hmmbuild 软件自制萝卜的隐马尔可夫模型(hidden markov model, HMM),之后利用 Hmmsearch 软件在象牙白萝卜中比对搜索 E 值小于 $1e^{-10}$ 的所有蛋白序列。并利用 NCBI 网站(<https://www.ncbi.nlm.nih.gov/>)的 Batch-CD-search 工具复筛含有完整 BELL、KNOX 结构域的萝卜 TALE 家族蛋白序列。利用 TBtools^[10] (<https://github.com/CJ-Chen/TBtools>) 进行保守结构域可视化。

1.3 萝卜 TALE 基因家族序列结构分析

根据鉴定得到的 TALE 基因家族成员及基因结构注释文件,利用 TBtools 中 gene structure view (advanced) 分析其基因结构。利用 MEME (<http://meme-suite.org/index.html>)对 TALE 基因家族氨基酸序列保守域和结构元件进行分析,基序数量寻找 10 个。

1.4 TALE 基因家族染色体定位与直系同源关系

根据鉴定出的象牙白萝卜 TALE 基因家族

成员信息及象牙白萝卜基因结构注释信息文件, 利用 TBtools 中的 gene location visualize from GTF/GFF 功能进行 *TALE* 基因家族的染色体定位。结合拟南芥和象牙白萝卜的基因结构注释信息文件及基因组文件, 使用 TBtools 中的 One step MCScanX 功能得到萝卜与拟南芥直系同源基因信息, 以‘36-2-R0N’表示序列所定位的萝卜染色体号, N(包括 0-9)为染色体编号, 以‘AT-ChrN’表示序列所定位的拟南芥染色体号, N(包括 0-5)为染色体编号。使用 TBtools 软件 dual systemy plot 功能绘制萝卜与拟南芥的直系同源基因图。

1.5 萝卜 *TALE* 基因家族启动子顺式作用元件分析

通过 PlantCARE 网站 (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) 的 ‘search for care’ 工具进行 *TALE* 基因起始密码子上游 2 000 bp 区域的顺式作用元件预测, 最后利用 TBtools 的 simple bioSequence viewer 功能对生成的结果进行可视化分析。

1.6 萝卜 *TALE* 家族基因表达特性分析

TALE 基因家族表达分析使用的转录组数据由本实验室测序完成。转录组数据包括‘象牙白’萝卜愈伤、肉质根、叶、茎、花、角果 6 个组织及‘心里美’萝卜肉质根芽期、破肚期、肉质根膨大前期、膨大盛期、成熟期 5 个发育时期^[11]。表达量经过 $\log_2(\text{RPKM})$ 标准化处理, 利用 TBtools 软件绘制热图^[12]。

1.7 萝卜 *TALE* 家族蛋白的理化性质与三维结构分析

蛋白长度、分子质量、等电点和亲水性等理化性质的分析通过在线工具 ExPASy (<https://web.expasy.org/protparam/>) 完成; 信号肽的判断分析通过在线工具 SignalP4.1Server (<http://www.cbs.dtu.dk/services/SignalP/>) 完成;

亚细胞定位预测通过在线工具 wolfsort (<https://wolfsort.hgc.jp/>) 完成; 蛋白三维结构建模分析通过在线工具 SWISS MODEL (<https://swissmodel.expasy.org/interactive>) 完成。

1.8 萝卜 *TALE* 基因家族系统进化分析

采用 MEGA-X 软件构建 *TALE* 基因家族系统进化树, 构建模型选择邻接法 (neighbor-joining), 利用 bootstrap method 和 pairwise deletion, 重复次数设置为 1 000。使用在线工具 iTOL (<https://itol.embl.de/>)^[13]美化系统进化树。

2 结果与分析

2.1 萝卜 *TALE* 转录因子家族成员的筛选与鉴定

经 Hmsearch 程序检索及 Batch-CD-search 工具复筛, 在象牙白萝卜 43 239 条氨基酸序列中找到 32 条符合条件的序列, 加上一条和拟南芥 KNATM 同源的氨基酸序列 Rsa10037940, 象牙白萝卜中共得到 33 条 *TALE* 氨基酸序列。除蛋白 Rsa10037940 外, 这些氨基酸序列均含有 1 个保守的 Homeobox-KN 结构域。33 条 *TALE* 氨基酸序列根据结构域的差异可以分为 BELL 和 KNOX 2 个亚家族。象牙白萝卜 BELL 亚族包含 21 个成员, 其中 17 个成员含有 POX、Homeobox-KN 结构域, 4 个成员含有 POX-superfamily、Homeobox-KN 结构域。象牙白萝卜 KNOX 亚族包含 12 个成员, 其中 7 个成员含有 KNOX2、KNOX1, Homeobox-KN、ELK 结构域, 2 个成员含有 KNOX2、KNOX1、Homeobox-KN 结构域, 成员 Rsa10036281 含有 KNOX2-superfamily、ELK、Homeobox-KN 结构域, 成员 Rsa10019161 除了含有 KNOX2、KNOX1、Homeobox-KN 3 种结构域之外, 还含有 Neurochodrin 结构域, 成员 Rsa10037940 仅含有 KNOX2-superfamily 结构域 (图 1)。

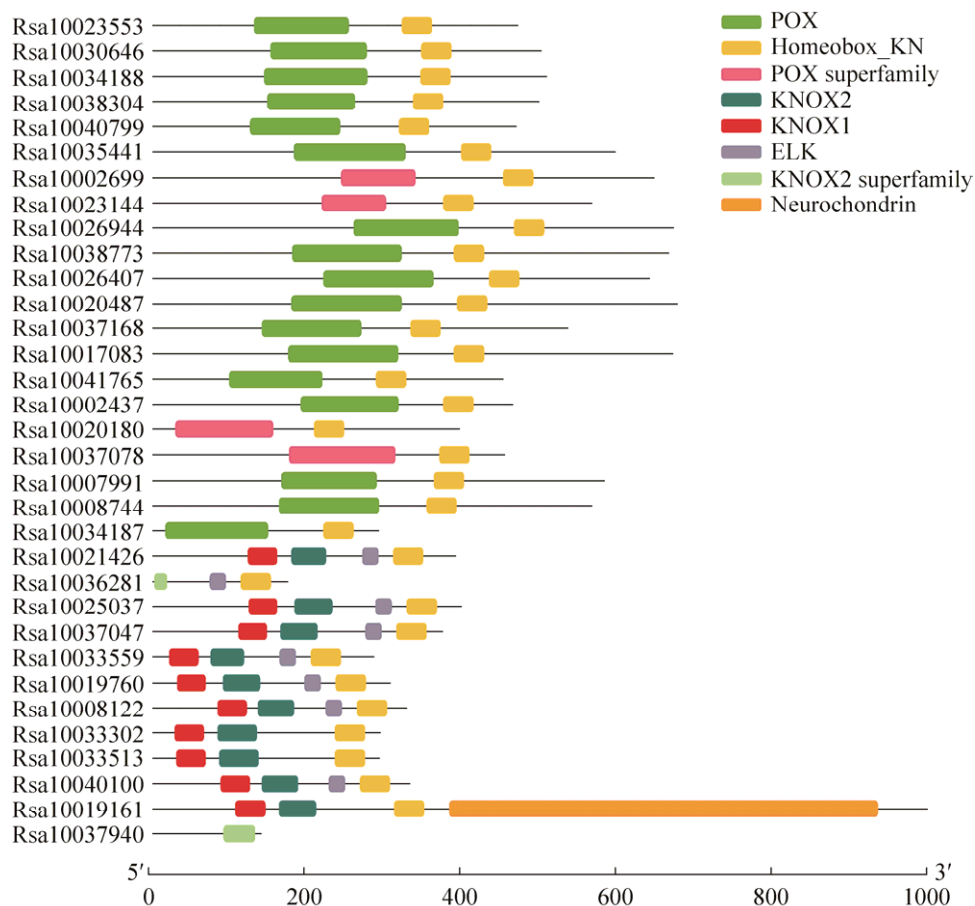


图1 萝卜 TALE 家族蛋白保守结构域

Figure 1 The conserved domain of the TALE family protein in radish.

2.2 萝卜 TALE 家族转录因子的蛋白结构分析

使用 MEGA-X 构建象牙白萝卜 TALE 家族氨基酸序列的系统进化树显示 (图 2A), 33 条萝卜 TALE 家族蛋白可以分为 8 个类群: KNOX 亚族分为 3 个类群, BELL 亚族分成 5 个类群。设置基序数量为 10, 利用 MEME 在线工具对萝卜 TALE 家族氨基酸序列进行分析发现, 不同的氨基酸序列包含的 motif 数量从 4-8 个不等 (图 2B)。除氨基酸序列 Rsa10037940 外, 其他家族成员均含有 motif1、motif3 保守基序, 说明 motif1、motif3 基序保守性最强, 对应结构域为 Homeobox-KN 结构域。除了共有基序 motif1、motif3 外, KNOX 亚家族 motif5 保守, POX 亚家族 motif4 保守, 各亚家族的独特保守

基序可能与各亚家族的不同功能相关。

2.3 萝卜 TALE 家族基因结构分析

在象牙白萝卜基因组中提取编码这些转录因子的基因, 发现 TALE 家族的基因序列结构差异较大 (图 2C), *Rsa10008122* 序列最长, 为 14 199 bp; *Rsa10036281* 序列最短, 仅 1 031 bp。各亚族之间内含子数目变化具有一定的规律性, BELL 亚族成员之间内含子个数在 2-4 之间, KNOX 亚族成员之间内含子个数在 3-5 之间。家族成员中除了 *Rsa10037940*、*Rsa10038304*、*Rsa10030646*、*Rsa10020180*、*Rsa10037078*、*Rsa10017083*、*Rsa10008122*、*Rsa10033513*、*Rsa10037168* 等序列含有 5'UTR 或 3'UTR 外, 其他成员均不含 5'UTR 和 3'UTR。序列含有

5'UTR 和 3'UTR 对于 mRNA 的稳定性有重要作用, 这些没有 UTR 的基因可能是没有注释出来, 而非真的没有 UTR。

2.4 萝卜全基因组 *TALE* 基因染色体定位以及与近缘物种的直系同源关系

TALE 家族基因在象牙白萝卜基因组 9 条染色体上均有分布, 从 Chr1 至 Chr9 分别有 3、7、

3、7、1、3、3、2 和 3 个基因 (图 3)。象牙白萝卜与拟南芥 *TALE* 基因家族成员共线性分析发现, 象牙白萝卜中 30 个 *TALE* 基因和拟南芥中 17 个 *TALE* 基因互为直系同源基因。在拟南芥 3 号染色体上没有直系同源基因分布。其中象牙白萝卜和拟南芥 *TALE* 家族基因中互为同源基因的对应关系见表 1 和图 4。

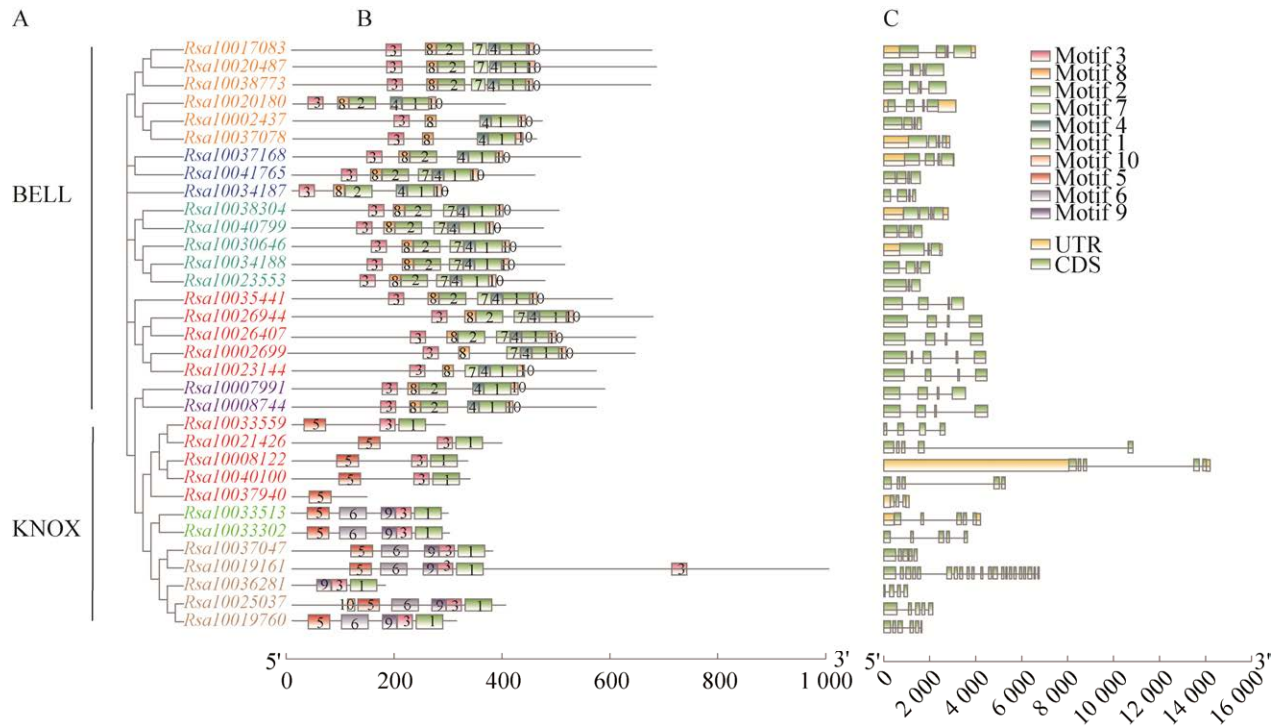


图 2 萝卜 *TALE* 家族成员基因结构和保守基序

Figure 2 Gene structure and conserved motif of the *TALE* family members in radish genome. (A) Phylogenetic tree. (B) Gene structure. (C) Conserved motifs.

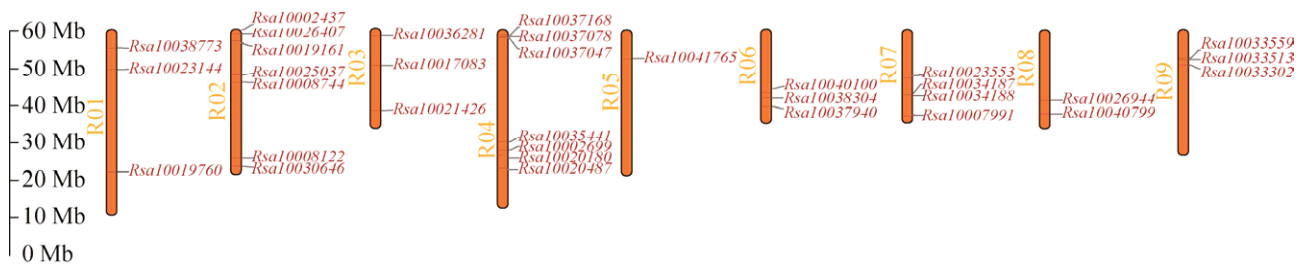


图 3 *TALE* 基因家族成员在萝卜染色体上的分布

Figure 3 Distribution of *TALE* family genes on radish chromosomes.

表 1 萝卜拟南芥同源基因共线性关系

Table 1 Co-linearity of the homologous genes in radish and *Arabidopsis*

AT-chromosome distribution	AT-gene ID		36-2-chromosome distribution	36-2-gene ID
AT-Chr1	<i>AT-AT1G14760.1</i>	==	36-2-R06	<i>36-2-Rsa10037940</i>
AT-Chr1	<i>AT-AT1G19700.1</i>	==	36-2-R08	<i>36-2-Rsa10040799</i>
AT-Chr1	<i>AT-AT1G62360.1</i>	==	36-2-R09	<i>36-2-Rsa10033559</i>
AT-Chr1	<i>AT-AT1G62990.1</i>	==	36-2-R09	<i>36-2-Rsa10033513</i>
			36-2-R09	<i>36-2-Rsa10033302</i>
AT-Chr1	<i>AT-AT1G70510.1</i>	==	36-2-R02	<i>36-2-Rsa10008122</i>
			36-2-R06	<i>36-2-Rsa10040100</i>
AT-Chr1	<i>AT-AT1G75410.1</i>	==	36-2-R02	<i>36-2-Rsa10030646</i>
			36-2-R07	<i>36-2-Rsa10023553</i>
			36-2-R07	<i>36-2-Rsa10034187</i>
AT-Chr2	<i>AT-AT2G16400.1</i>	==	36-2-R04	<i>36-2-Rsa10037168</i>
			36-2-R05	<i>36-2-Rsa10041765</i>
AT-Chr2	<i>AT-AT2G23760.1</i>	==	36-2-R01	<i>36-2-Rsa10023144</i>
			36-2-R04	<i>36-2-Rsa10002699</i>
AT-Chr2	<i>AT-AT2G27220.2</i>	==	36-2-R04	<i>36-2-Rsa10020180</i>
AT-Chr2	<i>AT-AT2G35940.1</i>	==	36-2-R01	<i>36-2-Rsa10038773</i>
			36-2-R03	<i>36-2-Rsa10017083</i>
			36-2-R04	<i>36-2-Rsa10020487</i>
AT-Chr4	<i>AT-AT4G08150.1</i>	==	36-2-R03	<i>36-2-Rsa10021426</i>
AT-Chr4	<i>AT-AT4G32040.1</i>	==	36-2-R01	<i>36-2-Rsa10019760</i>
			36-2-R02	<i>36-2-Rsa10019161</i>
			36-2-R04	<i>36-2-Rsa10037047</i>
AT-Chr4	<i>AT-AT4G32980.1</i>	==	36-2-R02	<i>36-2-Rsa10002437</i>
		==	36-2-R04	<i>36-2-Rsa10037078</i>
AT-Chr4	<i>AT-AT4G36870.1</i>	==	36-2-R02	<i>36-2-Rsa10026407</i>
AT-Chr5	<i>AT-AT5G02030.1</i>	==	36-2-R02	<i>36-2-Rsa10008744</i>
			36-2-R07	<i>36-2-Rsa10007991</i>
AT-Chr5	<i>AT-AT5G11060.1</i>	==	36-2-R02	<i>36-2-Rsa10025037</i>
AT-Chr5	<i>AT-AT5G41410.1</i>	==	36-2-R03	<i>36-2-Rsa10036281</i>
			36-2-R04	<i>36-2-Rsa10035441</i>

萝卜和拟南芥同为十字花科植物，亲缘关系较近，因此具有保守性较好的共线性区块。推测在进化过程中染色体水平上发生的融合、断裂，改变了萝卜和拟南芥基因的相对物理位置，甚至造成了 2 个物种染色体数目的变化，

但是家族基因所在位置的共线性排列依旧保守。除个别 *TALE* 家族基因外，萝卜 *TALE* 家族基因与拟南芥 *TALE* 家族基因有一对多的共线性关系，这一结果与萝卜在与拟南芥分化之后基因组发生过三倍化事件这一理论相吻合。

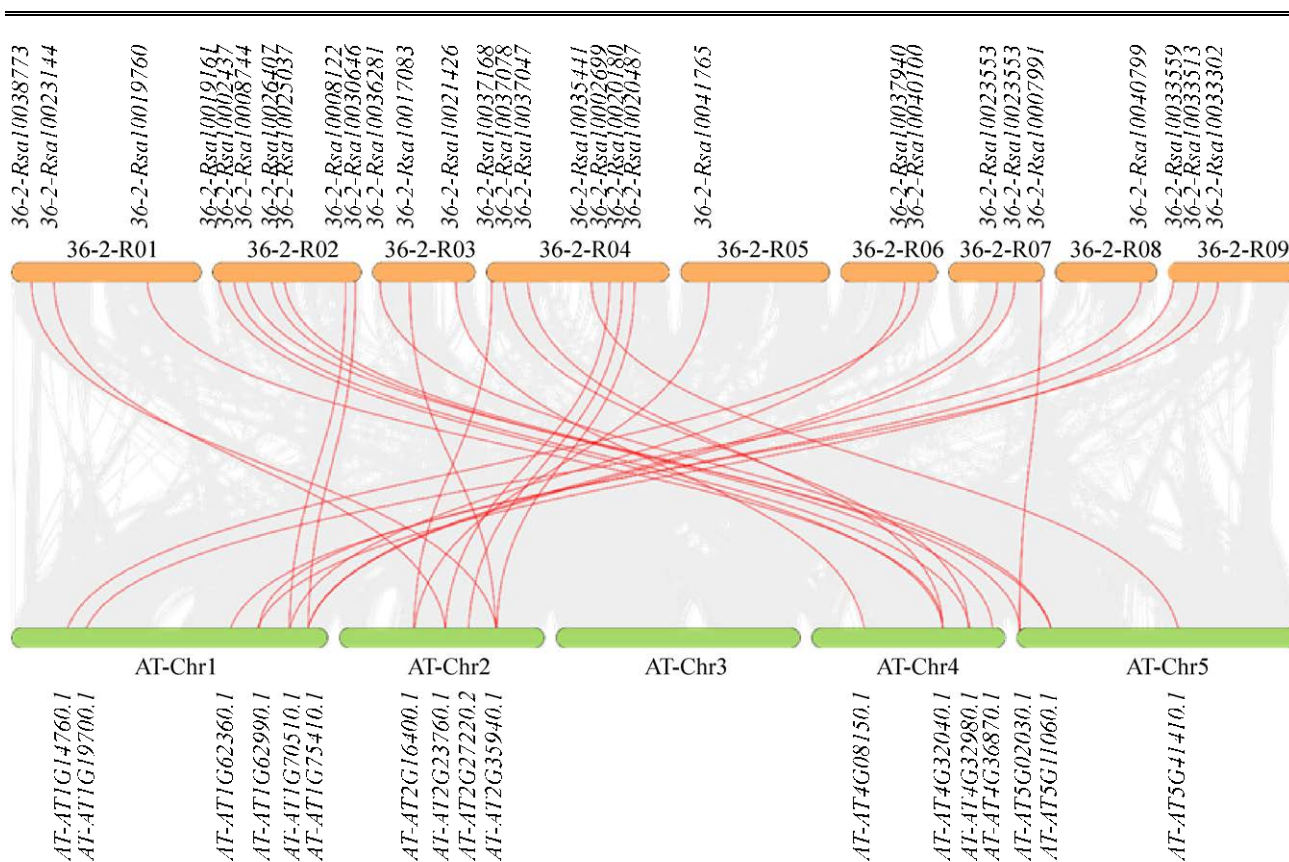


图4 萝卜拟南芥同源基因共线性关系 灰色线代表萝卜基因组与拟南芥基因组之间的共线性区块，红色线代表共线性区块中各 *TALE* 基因的位置

Figure 4 Co-linearity of homologous genes in radish and *Arabidopsis*. The gray line represents the collinear block between the radish genome and the *Arabidopsis* genome, and the red line represents the positions of each *TALE* gene in the collinear block.

2.5 萝卜 *TALE* 基因家族启动子特征分析

萝卜 *TALE* 基因家族启动子特征分析表明，其成员上游 2 000 bp 区域存在大量对赤霉素、生长素、低温、水杨酸、干旱诱导、脱落酸、茉莉酸甲酯响应的顺式元件 (图 5)。其中，*Rsa10040100* 启动子上只有 1 个脱落酸响应元件。66.7% 的成员包含茉莉酸甲酯响应元件，81.8% 的成员包含脱落酸响应元件，45.5% 的成员包含干旱诱导的 MYB 结合元件，60.6% 的成员包含生长素响应元件，30.3% 的成员包含低温响应元件，63.6% 的成员包含水杨酸响应元件，36.4% 的成员包含赤霉素响应元件。

2.6 萝卜 *TALE* 家族基因的表达特性分析

基于转录组数据对萝卜 *TALE* 基因家族成员的表达特性进行分析 (图 6)。结果显示，*TALE* 基因家族成员在萝卜不同组织 (肉质根、茎、愈伤、花、叶片、角果) 中表达量存在较大差异，*Rsa10026944* 和 *Rsa10026407* 在各组织中均高表达，*Rsa10002437* 在各组织中的表达量均较低；*Rsa10037047* 在茎中表达量较高，*Rsa10025037* 在愈伤组织中表达量较高；*KNOX* 亚族成员在不同组织间的表达量比较均一。

对萝卜 *TALE* 家族基因在‘心里美’萝卜肉质

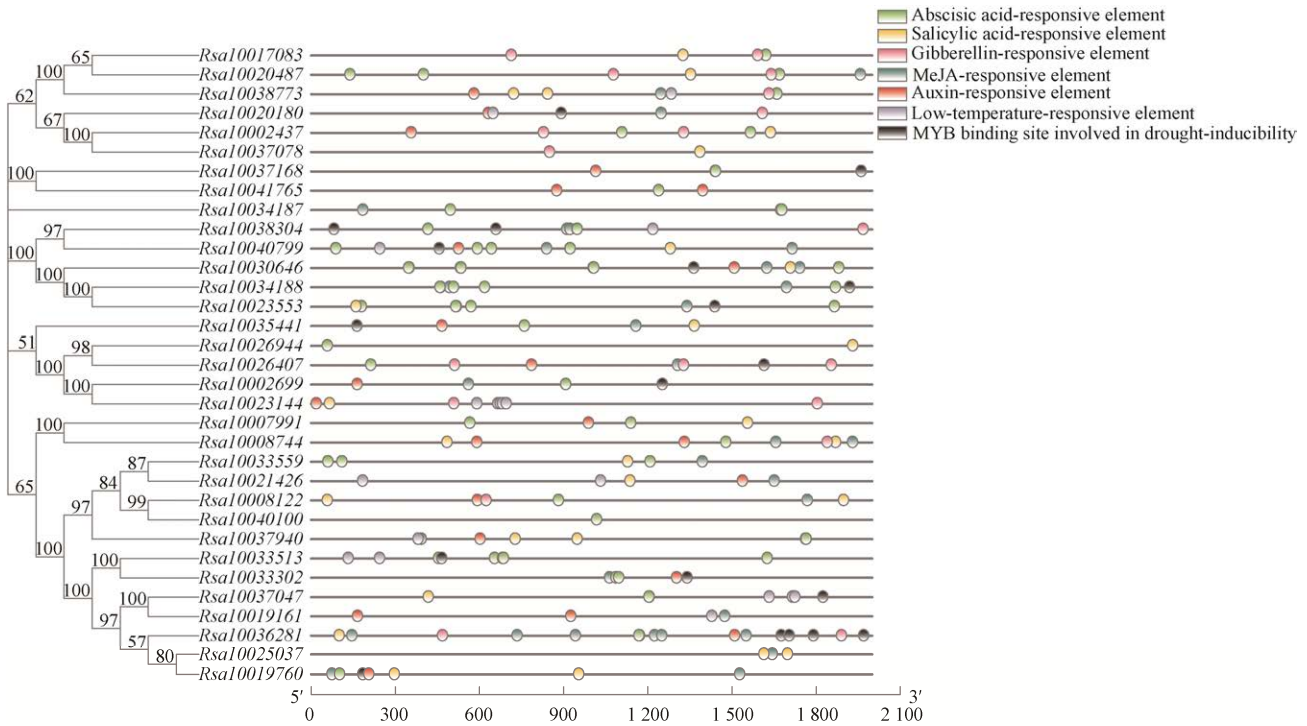


图 5 *TALE* 基因家族启动子顺式作用元件分布图

Figure 5 Cis-acting element distribution in the promoters of the *TALE* genes.

根 5 个发育时期中的表达量进行分析 (图 7), 发现 *Rsa10038773*、*Rsa10017083*、*Rsa10020487*、*Rsa10021426* 这 4 个基因在心理美材料的肉质根的 5 个不同发育时期中, 都具有较高的表达量。*Rsa10020180* 在心理美萝卜的 5 个生长时期呈现出表达量由低到高的表达趋势。

2.7 萝卜 *TALE* 基因家族蛋白的三维结构与理化性质分析

植物蛋白质的高级结构决定其生物学功能及活性大小。对象牙白萝卜 33 个 *TALE* 基因家族成员进行蛋白三维结构建模分析, 28 条氨基酸序列搜索到了相似度大于 30% 的模板, 其氨基酸三维结构表明不同亚型成员的蛋白 3D 结构具有高度的相似性。结构的主要成分为 3 个 α -螺旋。 β -折叠与 β -转角散布于整个蛋白序列中 (图 8)。另外 5 个蛋白质的结构可信度低, 本研究未采纳其预测的结构。如大小差异较大

的 *Rsa10037940* 和 *Rsa10019161* 等蛋白质。

利用生物信息学方式对 *TALE* 基因家族编码蛋白进行分析, 结果表明该家族蛋白长度变化范围为 139 aa (*Rsa10037940*) 到 995 aa (*Rsa10019161*); 蛋白分子量最大的为 110 116.7 Da (*Rsa10019161*), 最小的为 15 868.86 Da (*Rsa10037940*)。该家族蛋白除 *Rsa10002699*、*Rsa10007991*、*Rsa10020180* 以外均为酸性蛋白, 等电点小于 7。亲水性分析表明, 该家族蛋白均为亲水蛋白, 平均亲水性 (GRAVY) 均为负值。该家族成员均检测不到信号肽, 表明它们可能没有跨膜运输功能。亚细胞定位网站分析结果显示, 33 个编码蛋白均定位在细胞核, 这也与转录因子定位在细胞核调控基因表达的功能相吻合 (表 2)。

2.8 萝卜 *TALE* 转录因子家族进化关系

构建系统发育进化树对于基因功能分析十

分重要。利用氨基酸序列构建萝卜、拟南芥、水稻和胡杨的 NJ 系统发育树 (图 9), 结果显示在 4 个物种中 TALE 转录因子家族可以被划分为 Class I–VIII 共 8 组。其中 Class I、Class II、Class III 为 KNOX 亚族, Class IV、Class V、Class VI、Class VII、Class VIII 为 BELL 亚族。同一分组 TALE 转录因子家族氨基酸序列亲缘关系较近, 其中 Class I TALE 转录因子家族成员数量最多, 有 31 个; Class III TALE 转录因子家族成员数

量最少, 含有 7 个。由图可以看出萝卜和拟南芥的 TALE 氨基酸序列总是聚在一起, 同源性最高, 其次是胡杨, 水稻 TALE 氨基酸序列进化树距离最远, 有 1 个 KNOX 亚族成员和 3 个 BELL 亚族成员游离在双子叶植物的 8 个分支之外, 表明这些基因在单、双子叶植物中产生了独立的进化方向。TALE 转录因子家族内各亚族数量关系上, 拟南芥和萝卜 KNOX 亚族成员数量均少于 BELL 亚族成员数量, 但

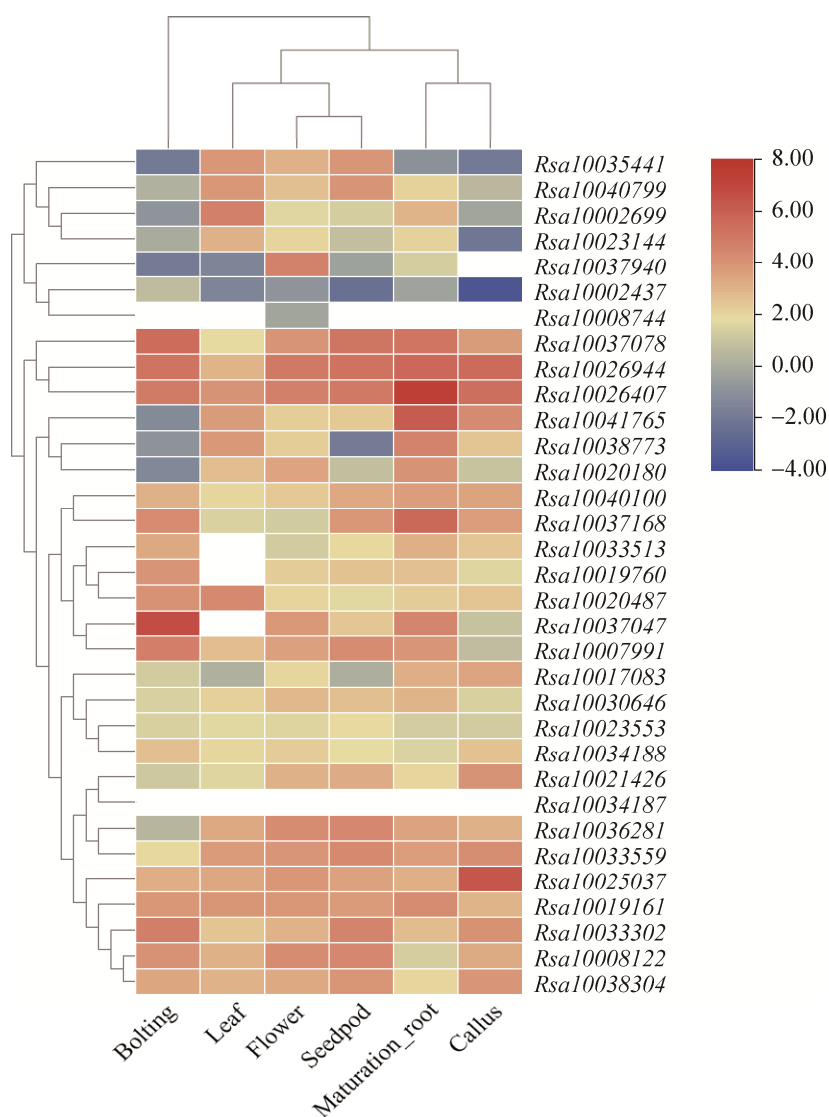


图 6 萝卜不同组织中 *TALE* 基因的差异表达

Figure 6 Expression of *TALE* genes in different tissues of radish.

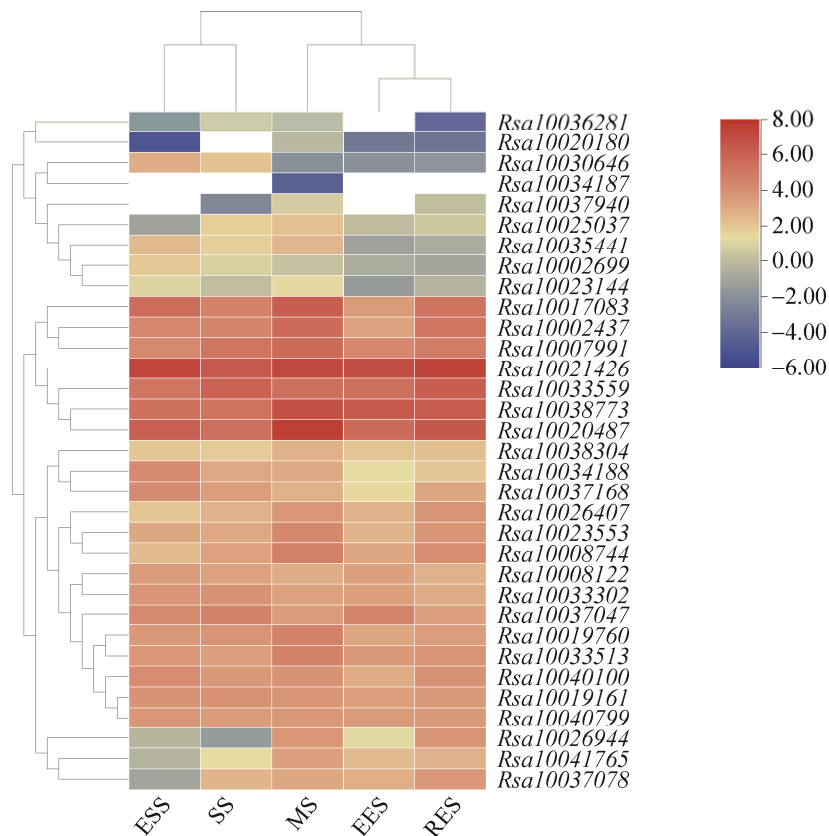


图 7 萝卜肉质根不同时期 *TALE* 基因的差异表达

Figure 7 Expression of *TALE* genes in radish roots at different stages.

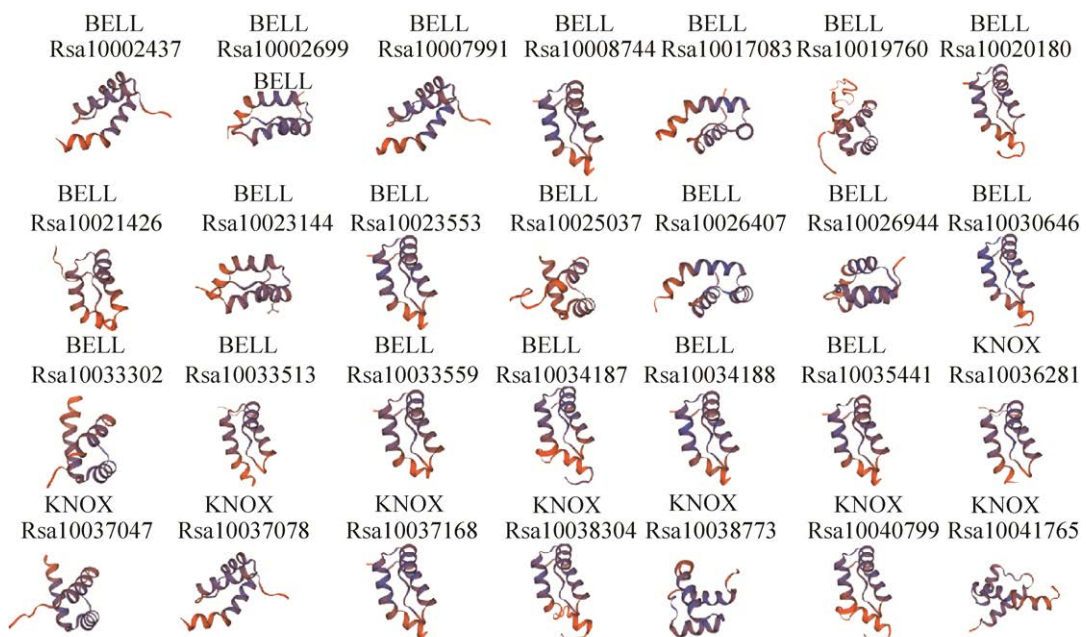


图 8 萝卜 *TALE* 家族成员三维结构

Figure 8 3D structure of *TALE* family members in radish.

表 2 萝卜 *TALE* 基因家族成员Table 2 Members of *TALE* gene family in radish

Genes ID	Domains	Length of aa	Molecular weight	Isoelectric point	Hydrophilic/hydrophobicity property	Sub-cellular localization	Homologue of At	AtTALE names
<i>Rsa10023553</i>	POX, Homeobox_KN	469	53 356.66	5.99	-0.94	nucl	<i>AT1G75410</i>	BLH3
<i>Rsa10030646</i>	POX, Homeobox_KN	499	56 885.91	6.73	-0.93	nucl	<i>AT1G75410</i>	BLH3
<i>Rsa10034188</i>	POX, Homeobox_KN	506	57 467.43	6.48	-0.92	nucl	<i>AT1G75410</i>	BLH3
<i>Rsa10038304</i>	POX, Homeobox_KN	496	56 562.60	6.14	-0.89	nucl	<i>AT1G19700</i>	BEL10
<i>Rsa10040799</i>	POX, Homeobox_KN	467	53 156.16	6.66	-0.86	nucl	<i>AT1G19700</i>	BEL10
<i>Rsa10035441</i>	POX, Homeobox_KN	594	67 209.75	6.67	-0.82	nucl	<i>AT5G41410</i>	BEL1
<i>Rsa10026944</i>	POX, Homeobox_KN	669	73 940.62	6.66	-0.76	nucl	<i>AT4G36870</i>	BLH2
<i>Rsa10038773</i>	POX, Homeobox_KN	663	73 247.31	6.87	-0.75	nucl	<i>AT2G35940</i>	BLH1
<i>Rsa10026407</i>	POX, Homeobox_KN	638	70 026.40	6.46	-0.71	nucl	<i>AT4G36870</i>	BLH2
<i>Rsa10020487</i>	POX, Homeobox_KN	674	73 376.10	6.44	-0.71	nucl	<i>AT2G35940</i>	BLH1
<i>Rsa10037168</i>	POX, Homeobox_KN	533	59 793.70	5.85	-0.70	nucl	<i>AT4G34610</i>	BLH6
<i>Rsa10017083</i>	POX, Homeobox_KN	668	73 084.08	6.63	-0.67	nucl	<i>AT2G35940</i>	BLH1
<i>Rsa10041765</i>	POX, Homeobox_KN	450	50 063.97	5.92	-0.59	nucl	<i>AT2G16400</i>	BLH7
<i>Rsa10002437</i>	POX, Homeobox_KN	462	52 582.17	6.23	-0.58	nucl	<i>AT4G32980</i>	ATH1
<i>Rsa10007991</i>	POX, Homeobox_KN	580	62 454.19	7.41	-0.48	nucl	<i>AT5G02030</i>	BLH9
<i>Rsa10008744</i>	POX, Homeobox_KN	564	60 716.34	6.91	-0.43	nucl	<i>AT5G02030</i>	BLH9
<i>Rsa10034187</i>	POX, Homeobox_KN	290	33 001.48	5.67	-0.34	nucl	<i>AT1G75430</i>	BLH11
<i>Rsa10037078</i>	POX superfamily, Homeobox_KN	452	51 406.73	5.99	-0.53	nucl	<i>AT4G32980</i>	ATH1
<i>Rsa10002699</i>	POX superfamily, Homeobox_KN	644	71 141.55	7.17	-0.80	nucl	<i>AT2G23760</i>	BLH4
<i>Rsa10023144</i>	POX superfamily, Homeobox_KN	564	62 102.09	6.15	-0.80	nucl	<i>AT2G23760</i>	BLH4
<i>Rsa10020180</i>	POX superfamily, Homeobox_KN	394	44 164.97	8.59	-0.54	nucl	<i>AT2G27220</i>	BLH5
<i>Rsa10021426</i>	KNOX1, KNOX2, ELK, Homeobox_KN	389	44 852.90	6.09	-1.06	nucl	<i>AT4G08150</i>	KNAT1
<i>Rsa10036281</i>	KNOX2 superfamily, ELK, Homeobox_KN	173	19 854.21	6.45	-0.98	nucl	<i>AT5G11060</i>	KNAT4
<i>Rsa10025037</i>	KNOX1, KNOX2, ELK, Homeobox_KN	396	44 754.73	6.29	-0.80	nucl	<i>AT5G11060</i>	KNAT4

(待续)

(续表 2)

Genes ID	Domains	Length of aa	Molecular weight	Isoelectric point	Hydrophilic/hydrophobicity property	Sub-cellular localization	Homologue of At	AtTALE names
<i>Rsa10037047</i>	KNOX1, KNOX2, ELK, Homeobox_KN	372	42 046.88	5.95	-0.78	nucl	<i>AT4G32040</i>	KNAT5
<i>Rsa10033559</i>	KNOX1, KNOX2, ELK, Homeobox_KN	284	32 277.55	6.13	-0.67	nucl	<i>AT1G62360</i>	STM
<i>Rsa10019760</i>	KNOX1, KNOX2, ELK, Homeobox_KN	305	33 869.99	5.69	-0.64	nucl	<i>AT5G25220</i>	KNAT3
<i>Rsa10008122</i>	KNOX1, KNOX2, ELK, Homeobox_KN	326	36 911.22	4.86	-0.64	nucl	<i>AT1G70510</i>	KNAT1
<i>Rsa10033302</i>	KNOX1, KNOX2, Homeobox_KN	292	32 865.06	6.10	-0.61	nucl	<i>AT1G62990</i>	KNAT7
<i>Rsa10033513</i>	KNOX1, KNOX2, Homeobox_KN	291	33 014.45	6.27	-0.56	nucl	<i>AT1G62990</i>	KNAT7
<i>Rsa10040100</i>	KNOX1, KNOX2, ELK, Homeobox_KN	330	36 914.56	4.89	-0.53	nucl	<i>AT1G23380</i>	KNAT6
<i>Rsa10019161</i>	KNOX1, KNOX2, Homeobox_KN, Neurochondrin	995	110 116.70	5.23	-0.14	nucl	<i>AT4G32040</i>	KNAT5
<i>Rsa10037940</i>	KNOX2 superfamily	139	15 868.86	4.82	-0.59	nucl	<i>AT1G14760</i>	KNATM

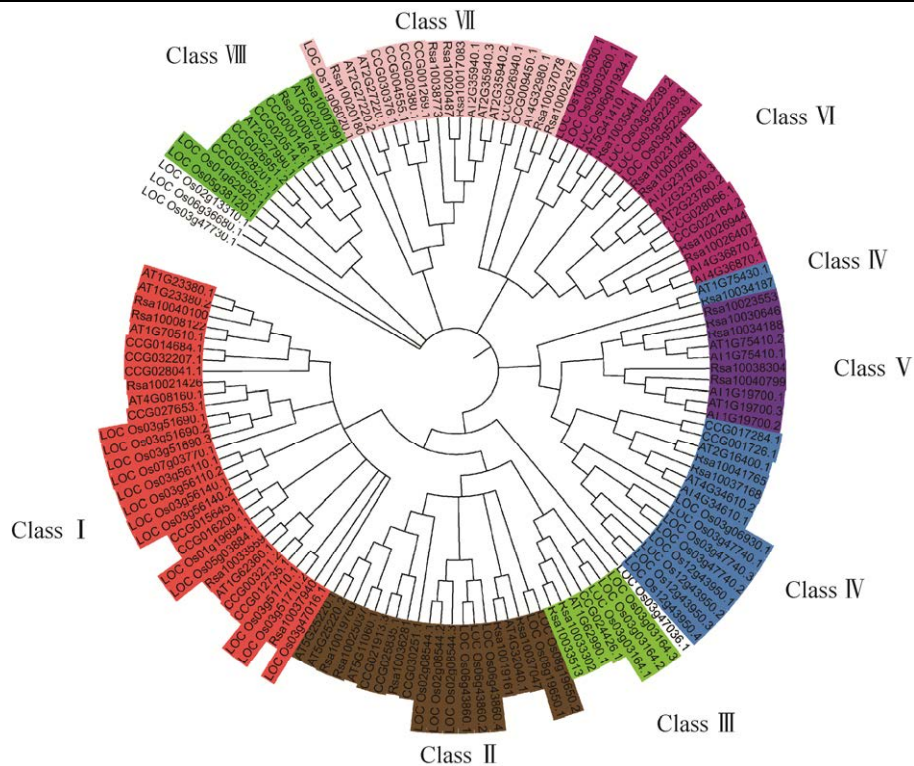


图 9 不同物种 TALE 转录因子家族成员进化树

Figure 9 Phylogenetic tree of TALE transcription factors from different species. At: *Arabidopsis thaliana*; Rsa: *Raphanus sativus*; CCG: *Populus euphratica*; LOC: *Oryza sativa*.

是胡杨和水稻 KNOX 亚族成员数量均多于 BELL 亚族成员数量,表明不同物种间 TALE 转录因子家族之间系统进化方向并不相同。

3 讨论

TALE 转录因子的结构域有 KNOX、BELL、Homeodomain 3 种,在调控植物生长发育、分生组织形成和维持器官形态发生方面发挥重要作用^[14]。KNOX 结构域对于顶端分生组织的调控和激素的合成途径有重要作用,研究发现,赤霉素合成途径中的关键酶基因 *GA20ox1* 可以被 KNOX 结构域抑制表达^[15]。BELL 蛋白与 KNOX 蛋白总是形成异二聚体发挥作用, BELL 蛋白调控分生组织和花的发育,但其功能冗余^[16]。

本研究通过对象牙白萝卜 TALE 基因家族的检索与筛选,在象牙白萝卜中共鉴定到了 33 个 TALE 基因家族成员。其基因差异表达分析发现 BELL 亚族的 *Rsa10026944*、*Rsa10026407*、*Rsa10037168*、*Rsa10041765* 主要在肉质根表达。*Rsa10026944*、*Rsa10026407* 在家族成员氨基酸序列进化关系中处于 Class VI, *Rsa10026407*、*Rsa10037168* 在家族成员氨基酸序列进化关系中处于 Class IV。*Rsa10026944*、*Rsa10026407* 与拟南芥对应的同源基因为 *BLH2*,其在种子粘液分泌细胞中显著表达,与种子萌发和幼苗早期发育有关^[17]。*Rsa10037168* 与拟南芥对应的同源基因为 *BLH6*,研究证明 *BLH6* 和 *BREVIPEDICELLUS* 之间存在相互作用的关系,共同调控根和茎的发育^[18]。而 KNOX 亚族的 *Rsa10037047*、*Rsa10025037* 主要在茎和愈伤组织中表达。*Rsa10037047*、*Rsa10025037* 与拟南芥对应的同源基因分别为 *KNAT4* 和 *KNAT5*,研究表明,*KNAT4* 和 *KNAT5* 在拟南芥根中具有细胞型特异性表达模式。此外,它们沿纵根轴和内根原基的表达具有不同的空间限制模

式^[19]。水稻胚胎发生过程中 KNOX 基因在茎间分生组织区域表达,表明这些基因参与了茎间分生形成的调节^[20]。目前关于 TALE 基因组织表达的研究表明:*BLH1* 与种子萌发和幼苗早期发育有关^[21],*BLH4* 调控叶形发育^[16],*BLH8* 在营养体向生殖体过渡过程中起着重要作用^[22],*BLH9* 调控花序茎的生长^[23],*KNAT6* 参与分生组织活性和器官分离^[24],*KNAT7* 与次生壁的形成有关^[25]。

萝卜和拟南芥的共线性分析发现,萝卜 TALE 家族 30 条基因序列与拟南芥中 17 条 TALE 基因序列具有共线性,多数萝卜 TALE 家族基因与拟南芥 TALE 家族基因有一对多的共线性关系,表明萝卜与拟南芥基因组分化后发生过基因家族复制事件,*Rsa10034188* 未在拟南芥中发现共线性基因。表明萝卜 TALE 家族成员在物种分化后进化出新的功能。

分析 TALE 家族成员的氨基酸序列进化关系、基因序列结构和氨基酸保守基序发现,系统发育进化树中关系紧密的成员外显子长度相似。含有相似基因结构的家族成员氨基酸保守基序也相似。TALE 转录因子家族蛋白 Homeobox-KN 结构域具有较强保守性,可作为后续功能验证的靶向序列。*Rsa10037940* 不含 Homeobox-KN 保守结构域,表明此基因在进化过程中结构域发生了丢失,基因功能也可能发生了改变。在萝卜 TALE 家族成员中 *Rsa10008122* 存在 1 个约 8 kb 的 5'UTR; *Rsa10019161* 除了含有 KNOX 保守结构域还多出来 1 个 Neurochondrin 结构域,经过和 WK10039 萝卜 (SAMN03084390) 基因组注释文件的比对,发现 *Rsa10008122* 和 *Rsa10019161* 的序列在萝卜 WK10039 基因组 (SAMN03084390) 中分别被注释成 2 个基因, *Rsa10008122* 中约 8 kb 的 5'UTR 在 WK10039 萝卜基因组中被单

独注释成一个基因, *Rsa10019161* 中多出来的结构域也被单独注释成一个基因, 表明 *Rsa10008122* 在象牙白萝卜基因组或 WK10039 萝卜基因组中有一个可能存在注释错误或拼接错误。

根据基因结构, 将萝卜 TALE 家族成员的氨基酸序列分成了 8 类, 其中 KNOX 亚族分为 3 类, BELL 亚族分为 5 类, 据报道, 拟南芥、棉花、石榴、杨树等植物的 TALE 家族氨基酸序列分类情况与此一致^[26], 表明植物中 TALE 基因家族序列保守性较强。拟南芥、萝卜、水稻、杨树 4 种植物的 TALE 蛋白系统进化分析中, BELL 和 KNOX 亚家族分支明显, 表明氨基酸序列上差异较大, 可能 2 个亚家族分化时间较早。萝卜和拟南芥在分化上存在一致性, 可能由于两者都属于十字花科, 亲缘关系较近。萝卜和水稻在分化上差异较大, 可能由于单双子叶植物在进化方向上有所不同。

启动子区域的顺式元件分析发现, TALE 基因启动子序列包含多个与激素反应和非生物胁迫相关的顺式元件。富含茉莉酸甲酯、脱落酸、水杨酸、赤霉素、生长素、低温诱导、干旱胁迫等响应元件, 表明萝卜 TALE 基因与逆境胁迫调控有关^[27]。这些结果与石榴^[26]、棉花^[8]等研究结果类似, 说明 TALE 基因启动子具有一定的保守性。

邓浪等发现信号肽与蛋白质定向转运功能有关^[28], 信号肽分析表明, 萝卜中 TALE 基因家族蛋白不参与细胞中蛋白质定向转运活动。蛋白理化性质分析表明, 萝卜 TALE 基因家族蛋白的分子质量和氨基酸数量差异较大, 表明 TALE 基因家族对于萝卜的调控方式可能更为复杂, 调控效果更加多样。

目前, TALE 基因家族已在多种植物中被鉴定出来, 但是关于萝卜 TALE 基因家族的鉴定

仍未见报道, 本研究利用生物信息学手段在象牙白萝卜中鉴定出了 33 个 TALE 基因家族成员, 研究了其理化性质及系统分化特点, 并结合转录组数据对基因表达特性也进行了分析, 萝卜 TALE 基因家族的鉴定及分析有助于后续对其成员的功能进行深入研究。

REFERENCES

- [1] Billeter M, Qian YQ, Otting G, et al. Determination of the nuclear magnetic resonance solution structure of an antennapedia homeodomain-DNA complex. *J Mol Biol*, 1993, 234(4): 1084-1097.
- [2] Jin J, Feng T, De-Chang Y, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*, 2017:D1040-D1045.
- [3] Magnani E, Hake S. KNOX lost the OX: the *Arabidopsis KNATM* gene defines a novel class of KNOX transcriptional regulators missing the homeodomain. *Plant Cell*, 2008, 20(4):875-887.
- [4] Meng L, Fan Z, Zhang Q, et al. *BEL1-LIKE HOMEODOMAIN 11* regulates chloroplast development and chlorophyll synthesis in tomato fruit. *Plant J*, 2018, 94(6): 1126-1140.
- [5] Zhao K, Zhang X, Cheng Z, et al. Comprehensive analysis of the three-amino-acid-loop-extension gene family and its tissue-differential expression in response to salt stress in poplar. *Plant Physiol Biochem*, 2019, 136: 1-12.
- [6] 张中荣, 吉雪花, 张海英, 等. 辣椒 TALE 转录因子的生物信息学分析. *分子植物育种*, 2020, 18(5): 1401-1408. Zhang ZR, Ji XH, Zhang HY, et al. Bioinformatics analysis of TALE transcription factors in pepper. *Mol Plant Breed*, 2020, 18(5): 1401-1408 (in Chinese).
- [7] 丘日光, 张天缘, 杨仕梅, 等. 百脉根 TALE 转录因子家族的鉴定与生物信息学分析. *植物遗传资源学报*, 2019, 20(2): 466-475. Qiu RG, Zhang TY, Yang SM, et al. Genome-wide identification and bioinformatics analysis of TALE transcription factor family in *Lotus japonicus*. *J Plant Genet Resour*, 2019, 20(2): 466-475 (in Chinese).
- [8] Ma Q, Wang N, Hao P, et al. Genome-wide identification and characterization of TALE superfamily genes in cotton reveals their functions in

- regulating secondary cell wall biosynthesis. *BMC Plant Biol*, 2019, 19(1): 432.
- [9] Zhang XH, Yue Z, Mei SY, et al. A *de novo* genome of a Chinese radish cultivar. *Hortic Plant J*, 2015, 1(3): 155-164.
- [10] Chen C, Chen H, Zhang Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*, 2020, 13(8): 1194-1202.
- [11] Sun Y, Wang J, Qiu Y, et al. Identification of 'Xinlimei' radish candidate genes associated with anthocyanin biosynthesis based on a transcriptome analysis. *Gene*, 2018, 657: 81-91.
- [12] 刘婷婷, 李晓曼, 张晓辉, 等. 萝卜全基因组中 *SPL* 基因家族成员的鉴定与分析. *植物遗传资源学报*, 2021, 22(4): 1145-1156.
Liu TT, Li XM, ZHANG XH, et al. Genome-wide identification and analysis of *SPL* gene family in radish. *J Plant Genet Resour*, 2021, 22(4): 1145-1156 (in Chinese).
- [13] Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*, 2016, 44(w1): W242-W245.
- [14] Shani E, Yanai O, Ori N. The role of hormones in shoot apical meristem function. *Curr Opin Plant Biol*, 2006, 9(5): 484-489.
- [15] Chen H, Banerjee AK, Hannapel DJ. The tandem complex of BEL and KNOX partners is required for transcriptional repression of *ga2ox1*. *Plant J*, 2004, 38(2): 276-284.
- [16] Kumar R, Kushalappa K, Godt D, et al. The *Arabidopsis* BEL1-LIKE HOMEODOMAIN proteins SAW1 and SAW2 act redundantly to regulate KNOX expression spatially in leaf margins. *Plant Cell*, 2007, 19(9): 2719-2735.
- [17] Xu Y, Wang Y, Wang X, et al. Transcription factors BLH2 and BLH4 regulate demethylesterification of homogalacturonan in seed mucilage. *Plant Physiol*, 2020, 183(1): 96-111.
- [18] Souček P, Hanáček P, Mazura P, et al. Interaction among BREVIPEDICELLUS, BLH6 and auxin in roots of *Arabidopsis thaliana*. *Russ J Plant Physiol*, 2017, 64(3): 386-397.
- [19] Truernit E, Siemerling KR, Hodge S, et al. A map of *KNAT* gene expression in the *Arabidopsis* root. *Plant Mol Biol*, 2006, 60(1): 1-20.
- [20] Postma-Haarsma AD, Verwoert II, Stronk OP, et al. 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 Mol Biol*, 1999, 39(2): 257-271.
- [21] Kim D, Cho YH, Ryu H, et al. BLH1 and KNAT3 modulate ABA responses during germination and early seedling development in *Arabidopsis*. *Plant J*, 2013, 75(5): 755-766.
- [22] Smith HMS, Campbell BC, Hake S. Competence to respond to floral inductive signals requires the homeobox genes *pennywise* and *pound-foolish*. *Curr Biol*, 2004, 14(9): 812-817.
- [23] Bhatt AM, EtcHELLS JP, Canales C, et al. VAAMANA—a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in *Arabidopsis*. *Gene*, 2004, 328: 103-111.
- [24] Belles-Boix E, Hamant O, Witiak SM, et al. KNAT6: An *Arabidopsis* homeobox gene involved in meristem activity and organ separation. *Plant Cell*, 2006, 18(8): 1900-1907.
- [25] Li E, Bhargava A, Qiang W, et al. The Class II KNOX gene KNAT7 negatively regulates secondary wall formation in *Arabidopsis* and is functionally conserved in *Populus*. *New Phytol*, 2012, 194(1): 102-115.
- [26] Wang YY, Zhao YJ, Yan M, et al. Genome-wide identification and expression analysis of *TALE* gene family in pomegranate (*Punica granatum* L.). *Agronomy*, 2020, 10(6): 829.
- [27] Michelmore RW, Meyers BC. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res*, 1998, 8(11): 1113-1130.
- [28] 邓浪, 沈兵琪, 王连春, 等. '红阳'猕猴桃全基因组 AP2/EREBP 转录因子生物信息学分析. *果树学报*, 2017, 34(7): 790-805.
Deng L, Shen BQ, Wang LC, et al. Genome-wide bioinformatics analysis of AP2/EREBP transcription factor in 'Hongyang' kiwifruit. *J Fruit Sci*, 2017, 34(7): 790-805 (in Chinese).

(本文责编 陈宏宇)