

百脉根 TALE 转录因子家族的鉴定与生物信息学分析

丘日光¹, 张天缘¹, 杨仕梅¹, 宋莉¹, 赵德刚^{1,2}

(¹ 贵州大学生命科学学院 / 农业生物工程研究院 / 山地植物资源保护与种质创新教育部重点实验室 /

贵州省农业生物工程重点实验室, 贵阳 550025; ² 贵州省农业科学院, 贵阳 550006)

摘要: 三胺酸环延伸 (TALE) 蛋白是一类在植物生长发育过程中调控分生组织分化的转录因子。本研究通过生物信息学手段从豆科模式植物百脉根 (*Lotus japonicus* (Regel) K.Larsen) 全基因组中筛选出分布于 6 条染色体上的 40 条 TALE 家族基因, 并对其保守结构域、基因结构、系统进化、在染色体上的分布、理化性质以及部分典型基因的组织表达差异等进行分析。根据结构域不同可将百脉根 TALE 家族分为 BELL 和 KNOX 两个亚族; 百脉根 TALE 家族在进化上较为保守, 分化上与大豆存在较大差异; 该家族基因有外显子 4~6 个, 氨基酸序列长度在 271~792 之间, 家族成员蛋白均为弱酸性蛋白。Real-time PCR 分析表明该家族基因表达与 motif 元件数之间存在相关性; BELL 亚族主要在顶芽表达, KNOX 亚族则主要在根组织中表达。研究结果为进一步克隆百脉根 TALE 基因和分析其功能奠定基础。

关键词: 百脉根; TALE 转录因子家族; 生物信息学; 系统进化; 基因表达

Genome-wide Identification and Bioinformatics Analysis of TALE Transcription Factor Family in *Lotus japonicus*

QIU Ri-guang¹, ZHANG Tian-yuan¹, YANG Shi-mei¹, SONG Li¹, ZHAO De-gang^{1,2}

(¹ College of Life Sciences and Institute of Agro-Bioengineering, Guizhou University/The Key Laboratory of Plant Resources

Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education) / Guizhou Key Lab of Agro-

Bioengineering, Guiyang 550025; ² Guizhou Academy of Agricultural Sciences, Guiyang 550006)

Abstract: TALE (three-amino acid-loop-extension) are essential transcription factors that regulate the differentiation of meristem in the plant growth and development. In this study, a total of 40 TALE family genes on six chromosomes were identified by bioinformatical analysis of the whole genome of the *Leguminosae* model plant *Lotus japonicus* (Regel) K.Larsen. Their conserved domain, gene structure, family evolution, chromosome distribution, physicochemical properties and expression differences were examined. As a result of the domain differences revealed that two major groups were divided into BELL and KNOX. Phylogenetic analysis found that the *Lotus japonicus* (Regel) K.Larsen TALE family gene sequences was conservative in evolution, while *Soybean* TALE gene was distinctly different with *Lotus japonicus* (Regel) K.Larsen in differentiation. The results of physicochemical properties showed they mainly possessed 4-6 exons encoding weak acid proteins of 271-792. The Real-time PCR result showed that TALE genes are diverse and exhibited different expression patterns in different tissues of *Lotus japonicus* (Regel) K.Larsen, related to the number of motif elements. BELL subfamily is highly

收稿日期: 2018-08-16 修回日期: 2018-09-14 网络出版日期: 2018-10-26

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20181025.1651.002.html>

第一作者研究方向为植物信号转导, E-mail: surivial@foxmail.com

通信作者: 宋莉, 研究方向为植物生物技术, E-mail: lpsssl@126.com

赵德刚, 研究方向为植物基因工程, E-mail: dgzhao@gzu.edu.cn

基金项目: 国家转基因生物新品种培育科技重大专项 (2016ZX08010003); 国家自然科学基金项目 (31660685); 贵州省科技计划项目 (黔科合基础 [2017] 1119)

Foundation project: The Genetically Modified Organisms Breeding Major Projects of China (2016ZX08010003), National Natural Science Foundation of China (31660685), Guizhou Science and Technology Major Project (20171119)

expressed in the terminal bud, while KNOX subfamily is Prominent in the root. This study lays a foundation for further study on the complicated biological functions of *Lotus japonicus* (Regel) K.Larsen TALE.

Key words: *Lotus japonicus* (Regel) K.Larsen; TALE transcription factor family; bioinformatics; phylogenetic analysis; gene expression

转录因子 (TF, Transcription factor) 是一类通过结合靶基因上游启动子区域、参与转录调控的蛋白质, 在植物生长发育、生理生化反应的信号转导和对外界环境的响应中发挥重要作用。三胺酸环延伸 (TALE, three-amino-acid-loop-extension) 蛋白是较早认识的一类动物特有 Homeobox 转录因子, 最近研究表明其也存在于拟南芥等植物中, 调控着分生组织的形成, 与器官形态的维持发生^[1]、器官位置^[2]、生殖过程、激素调控^[3]和信号转导^[4]等多个生物学过程密切相关。植物 TALE 转录因子家族中包含 BELL (BEL1-like homeodomain) 和 KNOX (KNOTTED-like homeodomain) 两种结构域, 因此将拟南芥 TALE 转录因子分为 BELL 亚族和 KNOX 亚族^[5], 同时这两个亚族具有特异性的互作关系^[6], 能够调控下游基因实现不同的生物学功能^[7]。

叶片是影响豆科牧草收获品质的主要因素^[8]。复叶物种可以通过多叶的特殊性状增加叶面积而增强光合作用, 进而增加产量^[9]。TALE 家族基因作为叶形成的重要调控因子, 调控复叶的起始、发育时期^[10], 在豆科牧草增产调控中有着重要价值。百脉根不仅是一种具有重要经济价值的豆科牧草, 也是植物分子生物学研究的重要模式植物, 其具有基因组结构简单^[11]和十分特殊的向基式复叶发育模式^[12]等特点。本研究通过对基因生物信息数据的挖掘和整理, 鉴定和分析了百脉根 TALE 家族基因的编码蛋白质性质以及各个成员间的亲缘关系, 为该基因家族功能研究及其分子育种利用和植物叶片发育及百脉根 TALE 及提供理论依据。

1 材料与方法

1.1 材料来源

百脉根 (*Lotus japonicus* (Regel) K.Larsen)、拟南芥 (*Arabidopsis thaliana* (L.) Heynh.)、水稻 (*Oryza sativa* subsp. *japonica* S.Kato) 和大豆 (*Glycine max* (L.) Merr.) TALE 转录因子信息来源于植物转录因子 (PlantTFDB, <http://planttfdb.cbi.pku.edu.cn/>) 数据库, 百脉根其他基因相关信息来源于百脉根

(miyakogusa.jp, <http://www.kazusa.or.jp/lotus/>) 数据库。

百脉根基因的组织表达差异分析材料于 2018 年 7 月 20 日取自贵州大学农业生物工程研究院试验田的里奥 (Leon) 百脉根。选取其幼苗期的根、茎、叶、顶芽作为材料, 液氮速冻后 -80 °C 冰箱保存。

1.2 分析方法

1.2.1 TALE 转录因子的鉴定与分类 使用 PlantTFDB4.0^[13] 中鉴定的百脉根 TALE 转录因子蛋白序列, 使用 Hmmbuild 软件构建百脉根的隐马尔可夫模型 (HMM, hidden markov model); 使用 Hmmssearch 软件筛选百脉根符合模型的所有蛋白序列 (E 值设为 1e-20)。通过 miyakogusa.jp 数据库获得与 ID 号对应的核酸序列和蛋白序列; 使用 Pfam (<http://pfam.xfam.org/>) 在线工具 (参数默认) 除去无 TALE 典型结构域的序列, 筛选百脉根 TALE 转录因子家族蛋白序列; 利用 TBtools (<https://github.com/CJ-Chen/TBtools>) 展示百脉根 TALE 转录因子家族蛋白结构域信息。

1.2.2 TALE 转录因子家族成员序列结构分析 对鉴定出的百脉根 TALE 家族成员, 利用在线工具 Gene Structure Display Sever (<http://gsds.cbi.pku.edu.cn/index.php>) 分析其基因结构; 使用 MEME (<http://meme-suite.org/>) 在线工具分析蛋白保守域和结构元件 (模体 motif 数量设置为 10)。

1.2.3 TALE 转录因子系统进化分析 使用 MEGA7.0 软件构建百脉根 TALE 转录因子家族邻接 (NJ, neighbour-joining) 树, 参数选择 Bootstrap method 和 pairwise deletion, 重复次数设置为 1000; 使用 FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) 软件进行系统发育树美化。使用 ClustalX2 软件对百脉根、拟南芥、水稻和大豆的 TALE 家族蛋白进行多序列比对; 使用 MEGA 7.0 软件绘制进化树, 采用 NJ 法, 选择 Bootstrap method 和 pairwise deletion 参数, 重复次数设置为 1000, 对 TALE 基因在不同植物中的进化关系进行分析; 使用网页工具 iTOL (<https://itol.embl.de/>)^[14] 绘制系统进化树。

1.2.4 百脉根 TALE 转录因子基因定位与直系同源关系 结合百脉根数据库的基因位置信息及染色

体长度,使用 PlantTFDB4.0 获得百脉根与大豆直系同源基因信息,以“ChrN”表示序列所定位的染色体号,N(包括 0~6)为染色体编号,以 Chr0 代表对所有无法定位到染色体的基因所虚构的一条染色体。使用 linux 软件 circos^[15]绘制百脉根与大豆的直系同源基因 circos 图。

1.2.5 TALE 蛋白的理化性质与三维结构分析 利用在线软件 ProtParam (<http://us.expasy.org/tools/protparam.html>),对百脉根 TALE 转录因子家族蛋白的蛋白长度、分子质量、等电点、总平均亲水性等理化性质进行分析;利用在线软件 SignalP4.1Server (<http://www.cbs.dtu.dk/services/SignalP/>)进行信号肽判断分析;利用在线软件 wolfsort (<https://wolfsort.hgc.jp/>)进行亚细胞定位预测;利用在线软件 SWISSMODEL (<https://swissmodel.expasy.org/interactive>)进行蛋白三维结构建模分析。

1.2.6 百脉根 TALE 家族基因的组织表达差异分析 使用 OMEGA 公司的 OMEGA Plant RNA Kit 提取百脉根总 RNA,并采用 ABI 公司的 ABI High-Capacity cDNA Reverse Transcriptase Kit 制备 cDNA,反应体系为:2.0 μ L 10 \times RT Buffer、0.8 μ L 25 \times dNTP Mix (100 mmol/L)、2.0 μ L 10 \times RT Random Primers、1.0 μ L MultiScribeTM Reverse Transcriptase、4.2 μ L Nuclease-free H₂O 配制的 2 \times RT Master Mix、10 μ L RNA。使用 ABI 公司的 ABI Power SYBR[®] Green PCR Master Mix RT-PCR 试剂盒进行 Real-time PCR 检测,采用百脉根(UB, ubiquitin)泛素基因作为内参基因对 motif 典型的 5 条 TALE 家族基因(表 1)进行组织差异表达检测。反应体系为:SYBR Premix Ex TaqTM (2 \times) 5.0 μ L、PCR Forward Primer (10 μ mol) 0.2 μ L、PCR Reverse Primer (10 μ mol) 0.2 μ L、cDNA 模板 1.0 μ L、ddH₂O 3.6 μ L。

表 1 qRT-PCR 所用引物序列

Table 1 Primers and sequences used for quantitative RT-PCR

名称 Name	上游引物 Forward primer (5'-3')	下游引物 Reverse primer (5'-3')
Lj0g3v0034399.1	TAAGGTTGACTCGGCGAAGA	CCTCTGCTCCACCTCATCAA
Lj0g3v0057839.1	AAAGGCCCTTCAGCAACTTG	TAAGCCTGTTTGCTAGCCA
Lj1g3v4752480.1	TGATGAGGTGGAGCAAAGGT	CTGCTCGAATTTGGCCTGTT
Lj3g3v1272300.1	CGCACAAACACTTCACCTCA	CGGTGGTGATGTCGTGAATC
Lj6g3v1088990.1	ATGTCCGTGTTTCATGCCATG	CACCATCCATGCTTCCTTCG
ljUB	GGGAGGTATGCAGATCTTTGTG	CCTCCAACCTAGAAATCCAC

2 结果与分析

2.1 百脉根 TALE 家族成员的筛选与鉴定

从 PlantTFDB 网站获得百脉根 TALE 家族蛋白序列共 31 条。通过 Hmmbuild 程序构建百脉根特有的 TALE 家族蛋白序列的隐马尔可夫模型矩阵(HMMER matrices)后,使用 Hmmssearch 程序检索百脉根 48105 条蛋白序列,在定义阈值为 1e-20 时,共得到符合条件的序列 41 条;这些序列经 Pfam 数据库搜索结构域后,去除一条不含 TALE 家族相关结构域的假阳性序列,共得到 40 条百脉根 TALE 基因家族成员蛋白序列。按结构域的不同,将百脉根 TALE 家族初步分为 BELL 和 KNOX 亚家族(图 1),两个亚族的 C 端均含有一个保守的 Homeobox 结构域。百脉根 BELL 亚族共有 28 条蛋白质,其中 21 条含有 BELL、Homeobox_KN、Homeobox 3 种结构域,其余的 7 条仅含有 BELL 结构域。KNOX 亚族共有 12 条蛋白质,其中 10 条含有 KNOX2、KNOX1、Homeobox_KN、Homeobox、ELK 5 种

结构域,其余两条存在不同程度的结构域缺失,如 Lj0g3v0250839.1 仅含有 Homeobox_KN 结构域。

2.2 百脉根 TALE 转录因子家族的结构分析

使用 MEGA7.0 构建的系统发育树表明(图 2),除了 Lj0g3v0250839.1 蛋白以外,其余 39 条百脉根 TALE 基因家族蛋白质分为两个类群,与 BELL 和 KNOX 两种亚型的不同结构域的分布具有一致性,即蛋白结构域相同的,以及同一个基因的扩增或者复制的蛋白质亲缘关系较近,而 Lj0g3v0250839.1 因为结构域缺失导致其单独形成一枝。百脉根 TALE 转录因子家族各成员的基因结构差异较大,以 Lj4g3v0911380.1 序列最长,为 9411 bp,而 Lj1g3v1785910.1 最短,仅 4458 bp;各成员的内含子数目差异也较大,最多的可含有 6 条内含子,成员 Lj0g3v0250839.1 则没有内含子;家族所有成员序列均含有 5' UTR,对于 mRNA 的稳定性有重要作用,而 Lj0g3v0057839.1、Lj0g3v0057839.2 两条序列含有 3' UTR 的结构,这有助于结合 microRNA,进而调控转录因子蛋白的表达量。

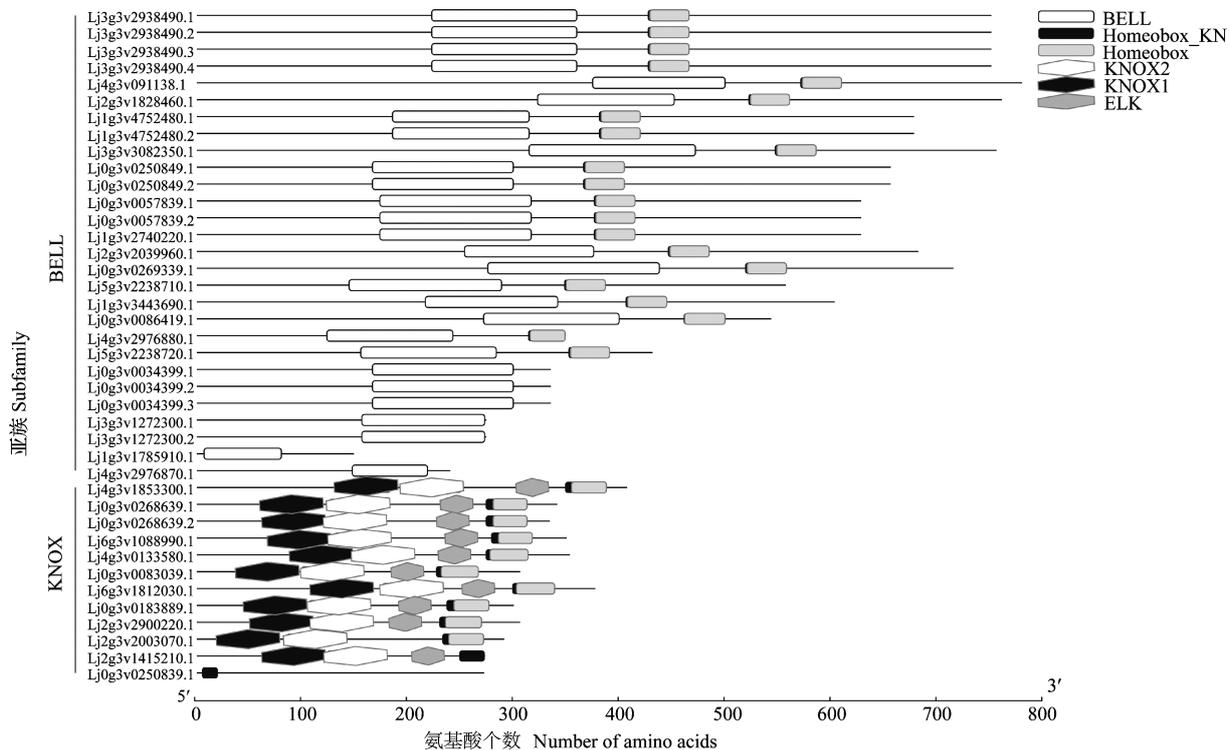
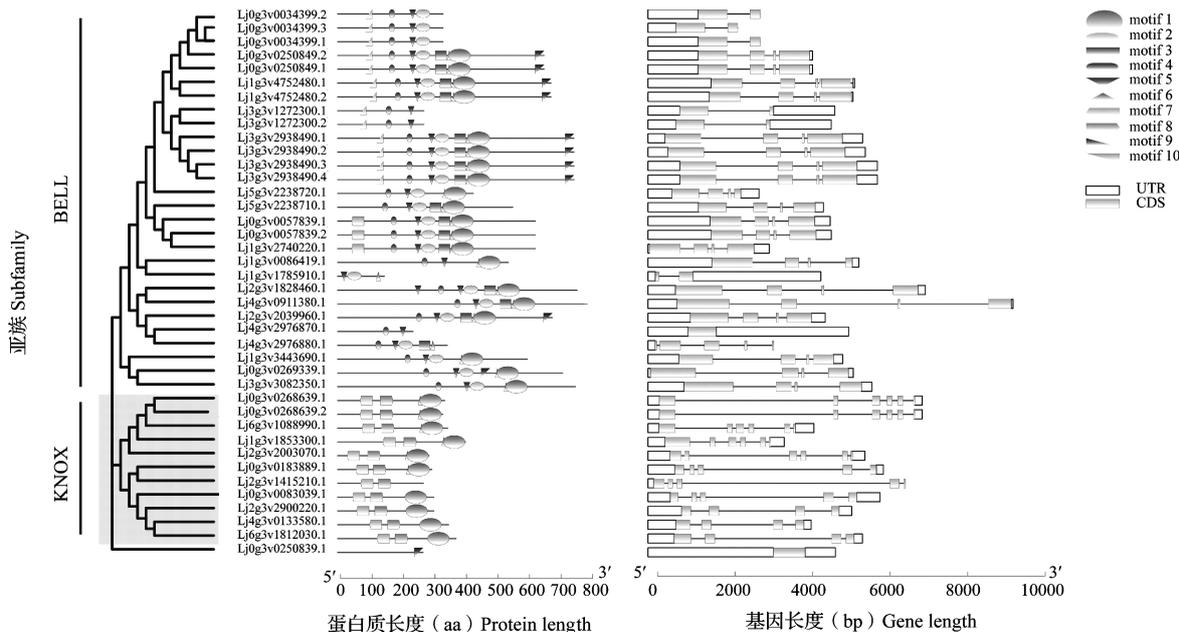


图 1 百脉根 TALE 转录因子家族的蛋白质结构域

Fig.1 protein domain of *Lotus japonicus* (Regel) K.Larsen TALE transcription factor family



图中从左至右分别为百脉根 TALE 转录因子家族的进化关系、motif 组成和基因结构

In the figure, from left to right, are the evolutionary relationship, motif and gene structure of the *Lotus japonicus* TALE transcription factor family

图 2 百脉根 TALE 转录因子家族的基因序列结构和 motif 组成

Fig.2 Gene structure and motif of *Lotus japonicus* (Regel) K.Larsen TALE transcription factor family

利用 MEME 在线工具对百脉根 TALE 基因家族成员进行分析发现,其蛋白序列具有 10 种 motif (图 3),且存在不同程度的缺失,含有的 motif 数量从 1~8 条不等,但与亚家族划分和结构域分布相一致;其中, motif2、motif3、motif4、motif5、motif9、motif10

属于 BELL 结构域, motif7、motif8 属于 KNOX 结构域, motif1、motif6 属于 Homeobox 结构域;这些 motif 的期望值均较低,长度从 15~50 个氨基酸不等;而不同的蛋白序列包含的 motif 数量也有较大差异,从 4~8 个不等。

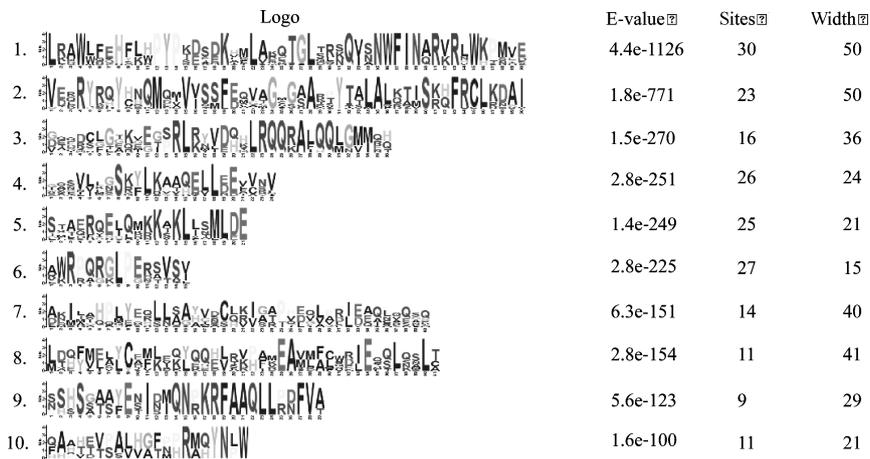


图 3 百脉根 TALE 转录因子家族蛋白的 motif 类型

Fig.3 Protein motif variety of *Lotus japonicus* (Regel) K.Larsen TALE transcription factor family

2.3 百脉根 TALE 转录因子家族进化关系

目前对百脉根 TALE 基因家族的了解较少,其进化关系的分析有助于对基因功能的揭示。利用百脉根(40 条)、拟南芥(33 条)、大豆(133 条)和

水稻(45 条) TALE 家族蛋白序列构建的 NJ 系统发育树表明,植物 TALE 家族成员可以被划分为 6 类原始的基因簇(图 4)。系统进化枝(clade)尖端较短的枝长表明,聚类的植物 TALE 家族成员

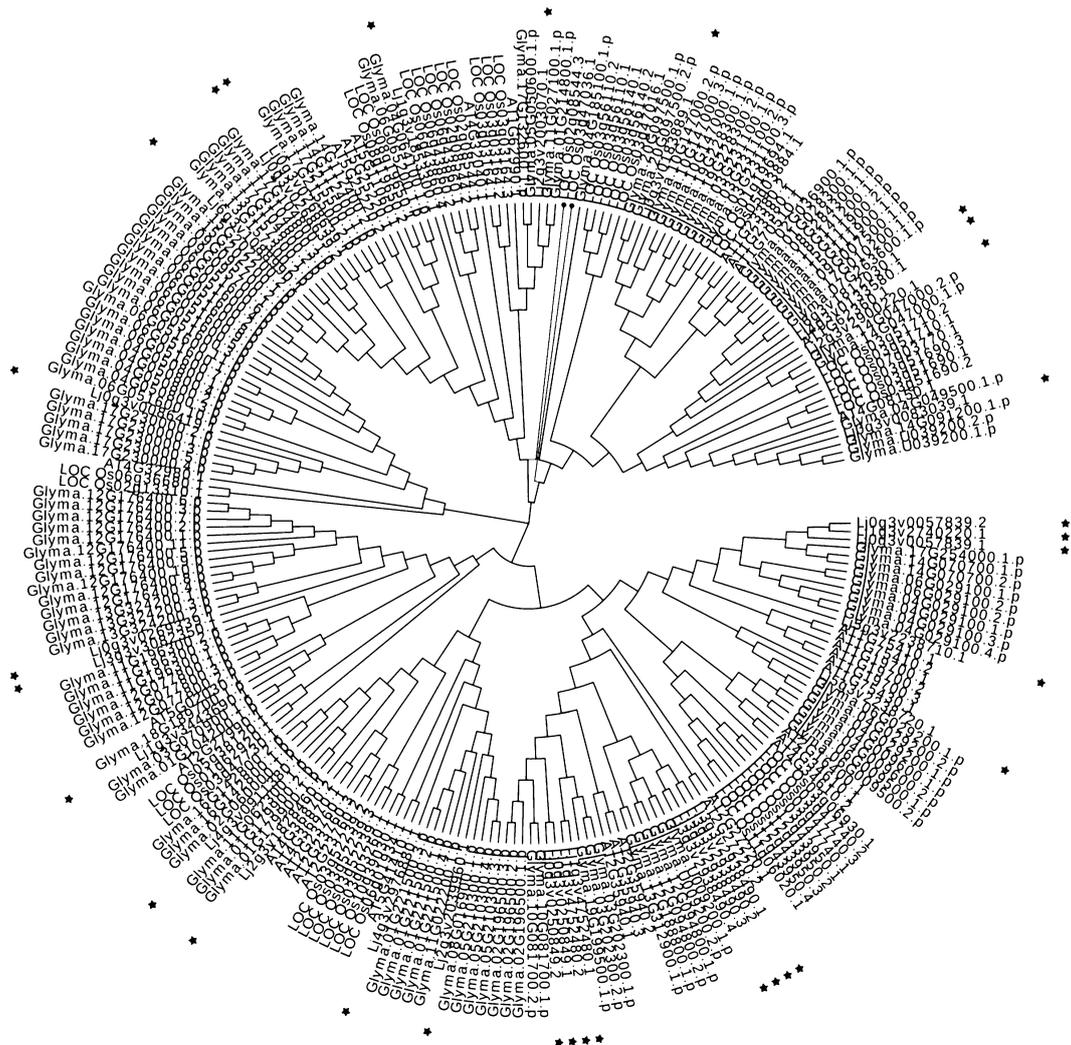


图 4 百脉根、拟南芥、大豆、水稻蛋白的系统进化分析

Fig.4 Phylogenetic analysis of *Lotus japonicus* (Regel) K.Larsen, *Arabidopsis thaliana* (L.) Heynh., *Glycine max* (L.) Merr., *Oryza sativa* subsp. *Japonica* S.Kato TALE transcription factor family members

具有较强的氨基酸保守性,各家族成员之间进化关系紧密;在每个进化枝中,具有多个来自相同物种的分枝,可能是基因复制事件引起的,而缺少某些物种的代表可能与该家族发生了基因丢失事件有关;同时每个进化枝中均包含不同物种成员,如百脉根中的 Lj5g3v2238720.1、Lj4g3v2976880.1、Lj2g3v2039960.1、Lj1g3v3443690.1、Lj0g3v0086419.1、Lj0g3v0083039.1、Lj2g3v2003070.1 几条基因和拟南芥在进化关系上较为接近,可能它们在进化上较为保守。4种植物中的 BELL 和 KNOX 亚家族分支非常明显,表明两个亚家族的分化发生在单双子叶分化之前。大豆的 TALE 家族成员数量与其他 3 个物种差异较大,可能是大豆本身发生了基因复制事件。

2.4 百脉根全基因组 TALE 转录因子定位与直系同源关系

百脉根与大豆 TALE 家族成员分化关系的 circos 图(图 5)表明,有 24 条百脉根基因和 63 条大豆基因属于直系同源基因,这些直系同源基因在各条染色体上均有分布;其中,百脉根除去无法确定位置的 7 个成员外(分布在 Chr0 上),其他成员的分布较为均匀,在 Chr1~Chr6 上分别有 4、4、2、3、2、2 条基因;直系同源基因在大豆中的分布不均匀,在 Chr1(5 条)、Chr4(7 条)、Chr6(5 条)、Chr17(6 条)上分布较多,其他染色体上分布较少,数量在 1~4 条之间。

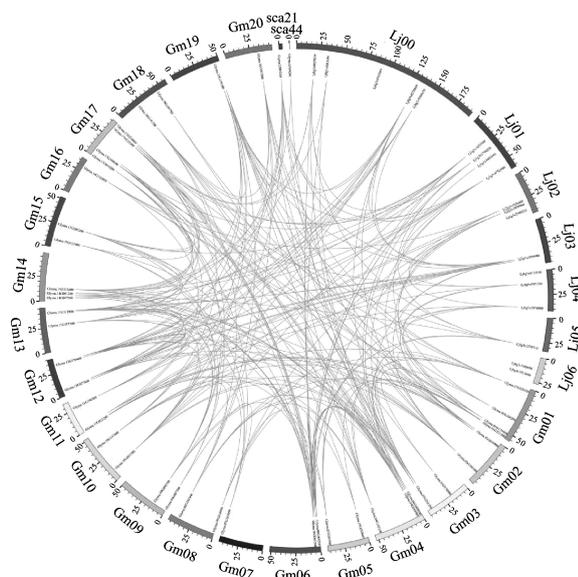


图 5 百脉根与大豆直系同源基因的 circos 图

Fig.5 Orthologous genes circos of *Lotus japonicus* (Regel) K.Larsen and *Glycine max* (L.) Merr. TALE transcription factor family

2.5 TALE 蛋白的理化性质与三维结构分析

对 TALE 转录因子家族编码蛋白的分析表明(表 2),各亚族之间的蛋白质理化性质存在一定差异。在组成蛋白长度的氨基酸数量上,从 271aa (Ljog3v0250839.1) 到 792aa (Lj4g3v0911380.1) 不等;蛋白分子量也存在差异,最大的为 11943 kD (Lj4g3v0911380.1),最小的为 4614 kD (Lj2g3v2003070.1)。除了 Lj4g3v2976880.1、Lj1g3v1785910.1 和 Lj1g3v3443690.1 以外,所有 TALE 蛋白的等电点均小于 7,表示多数 TALE 家族基因可能编码弱酸性蛋白,在酸性亚细胞环境中发挥作用。40 条百脉根 TALE 蛋白的总平均亲水性 (GRAVY) 均为负值,表明 TALE 家族蛋白均为亲水性蛋白。BELL 亚族中以 Lj2g3v2039960.1 的亲水性最大,为 -0.825; KNOX 亚族中则以 Lj0g3v0268639.2 亲水性最大,为 -0.882。不同亚类间的 TALE 蛋白的理化特性变化均较大,具有不同特性;在 3 个亚类中, KNOX domain 亚类蛋白的理化性质差异最小。信号肽分析表明,TALE 家族成员均不含有信号肽,难以在细胞中进行跨膜运输。亚细胞定位结果显示,百脉根 TALE 家族定位在细胞核,这也与其作为转录因子的功能相吻合。

植物蛋白质的高级结构决定其生物学功能及活性大小。蛋白二级结构预测分析表明,在 40 条百脉根 TALE 家族成员中,28 条蛋白的氨基酸序列以 α -螺旋为主要组成成分,其余则以无规则卷曲为主要组成成分, β -折叠与 β -转角散布于整个蛋白序列中。选取 2 个亚型中具有代表性的序列进行三级结构预测, BELL 亚型以 Lj3g3v2938490.1 为代表, KNOX 亚型以 Lj6g3v1088990.1 为代表,结果表明不同亚型之间在 3D 结构上具有高度的相似性(图 6),但 BELL 亚型成员的主要元件为 α -螺旋,与 KNOX 亚型有所区别。

2.6 百脉根 TALE 基因组织表达差异分析

根据 TALE 基因序列 motif 组成元件的差异,选取了 5 个典型的百脉根 TALE 转录因子家族基因,用 RT-PCR 对其在百脉根的根、茎、叶、顶芽中的表达进行分析。结果显示(图 7), motif 元件数越多,基因表达量越高;5 个 TALE 基因在以上 4 种组织中均有表达,但以顶芽和根的表达量最高,存在组织表达差异。BELL 亚族的 Lj0g3v0034399.1、Lj1g3v4752480.1 主要的表达部位是顶芽,而 BELL 亚族 Lj3g3v1272300.1、Lj0g3v0057839.1 和 KNOX 亚族 Lj6g3v1088990.1 3 条基因在根的表达量最高。这可能与其分属于 TALE 家族的两个亚族有关。

表 2 百脉根 TALE 转录因子蛋白理化性质

Table 2 protein physical and chemical properties of *Lotus japonicus* (Regel) K.Larsen TALE transcription factor

亚族 Subfamily	编号 Number	蛋白长度 (aa) Protein length	分子量 (kD) Molecular mass	等电点 Isoelectric point	总平均亲水性 Grand average of hydropathy	信号肽 Signal peptide	亚细胞定位 Subcellular localization	
BELL	Lj0g3v0034399.1	334	35559	6.84	-0.450	否	nucl	
	Lj0g3v0034399.2	334	35559	6.84	-0.450	否	nucl	
	Lj0g3v0034399.3	334	35559	6.84	-0.450	否	nucl	
	Lj0g3v0057839.1	627	9599	5.83	-0.685	否	nucl	
	Lj0g3v0057839.2	627	9599	5.83	-0.685	否	nucl	
	Lj1g3v2740220.1	627	9599	5.83	-0.685	否	nucl	
	Lj0g3v0250849.1	665	9881	6.57	-0.609	否	nucl	
	Lj0g3v0250849.2	665	9881	6.57	-0.609	否	nucl	
	Lj1g3v4752480.1	677	10191	6.42	-0.708	否	nucl	
	Lj1g3v4752480.2	677	10191	6.42	-0.708	否	nucl	
	Lj2g3v1828460.1	760	11314	6.56	-0.690	否	nucl	
	Lj5g3v2238710.1	556	8631	5.67	-0.653	否	nucl	
	Lj3g3v2938490.1	750	11291	6.39	-0.704	否	nucl	
	Lj3g3v2938490.2	750	11291	6.39	-0.704	否	nucl	
	Lj3g3v2938490.3	750	11291	6.39	-0.704	否	nucl	
	Lj3g3v2938490.4	750	11291	6.39	-0.704	否	nucl	
	Lj4g3v0911380.1	792	11943	6.18	-0.739	否	nucl	
	Lj3g3v3082350.1	755	11416	6.78	-0.683	否	nucl	
	Lj5g3v2238720.1	430	6599	6.25	-0.424	否	nucl	
	Lj4g3v2976880.1	348	5562	9.06	-0.778	否	nucl	
	Lj0g3v0269339.1	714	10972	6.49	-0.582	否	nucl	
	Lj2g3v2039960.1	681	10535	6.25	-0.825	否	nucl	
	Lj1g3v3443690.1	602	9188	7.83	-0.545	否	nucl	
	Lj0g3v0086419.1	542	8287	5.56	-0.415	否	nucl	
	Lj3g3v1272300.1	273	29337	6.45	-0.606	否	nucl	
	Lj3g3v1272300.2	273	29337	6.45	-0.606	否	nucl	
	Lj1g3v1785910.1	148	17285	7.14	-0.605	否	nucl	
	Lj4g3v2976870.1	239	27342	6.26	-0.615	否	nucl	
	KNOX	Lj6g3v1812030.1	376	5767	6.05	-0.707	否	nucl
		Lj4g3v0133580.1	352	5452	5.95	-0.603	否	nucl
Lj2g3v2900220.1		305	4854	6.67	-0.716	否	nucl	
Lj0g3v0183889.1		299	4733	5.01	-0.673	否	nucl	
Lj0g3v0083039.1		305	4887	6.43	-0.861	否	nucl	
Lj1g3v1853300.1		406	6268	6.15	-0.775	否	nucl	
Lj6g3v1088990.1		349	5481	6.19	-0.748	否	nucl	
Lj0g3v0268639.2		333	5328	6.03	-0.882	否	nucl	
Lj0g3v0268639.1		340	5399	5.77	-0.839	否	nucl	
Lj2g3v2003070.1		290	4614	6.31	-0.628	否	nucl	
Lj2g3v1415210.1		272	30420	4.82	-0.577	否	nucl	
Lj0g3v0250839.1		271	30177	5.36	-0.687	否	nucl	

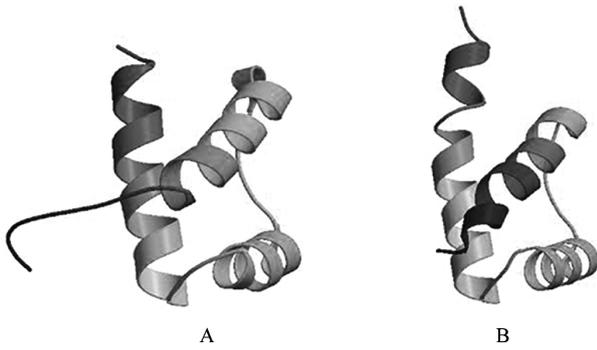


图 6 百脉根 TALE 家族中 KNOX 亚型 (A) 和 BELL 亚型 (B) 的 3D 结构

Fig.6 Predicted 3D structure of the KNOX subtype (A) and BELL subtype (B) of the *Lotus japonicus* (Regel) K.Larsen TALE transcription factor family

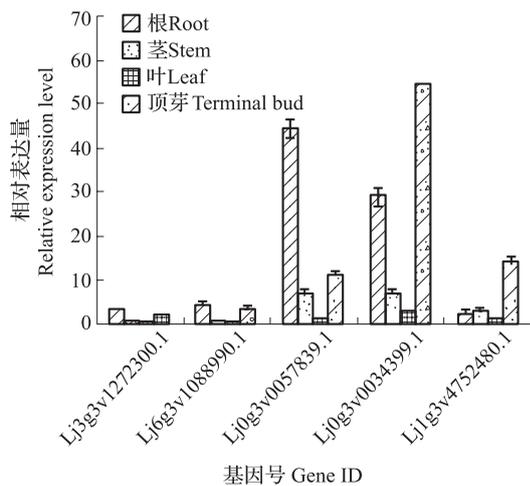


图 7 百脉根 TALE 基因在不同组织中的表达情况

Fig.7 Expression level of TALE genes in different tissues from *Lotus japonicus* (Regel) K.Larsen

3 讨论

转录因子作为一类参与发育调控的蛋白质,在植物生长发育过程中发挥关键作用^[16]。TALE 转录因子是一类调控植物生长发育、分生组织形成和维持以及器官形态发生等的重要转录因子。TALE 转录因子的结构域包括 KNOX、ELK 和 Homeobox 3 种。KNOX 结构域常存在于植物顶端分生组织中,与分生组织的分化和信号转导有关,如可以抑制 GA 途径中关键酶基因 *ga20ox1* 的表达^[17]; ELK 结构域则是一个两性螺旋结构,对于蛋白质相互作用有重要作用; Homeobox 结构域则用于和靶基因的结合。拟南芥 TALE 家族基因 *VAMANA* 能影响拟南芥花序与叶片生长,形成矮秆性状^[18]; 大麦 KNOX 亚族成员由于不同的组织细胞极性调控分生组织产生不同的形态效应^[19]; 土豆中 TALE 家族

成员中的 *POTHI* 基因,可作为发育调节的移动信号,调节激素水平,进而加快块茎的形成^[20]。但目前为止,人们对 TALE 转录因子在百脉根发育中的调控了解较少。

本文通过建立百脉根特有的 TALE 转录因子家族 HMM 模型,对百脉根全基因组数据的检索与筛选,共鉴定出 40 条百脉根 TALE 转录因子家族成员,依据结构域特点将其划分为 BELL 和 KNOX 两个亚族。40 条 TALE 转录因子中仅有 2 条序列含有 3' UTR,这种结构作为 miRNA 可能的结合位点^[21],体现了 TALE 转录因子家族对于下游基因具有更加丰富的调控特性。TALE 转录因子家族蛋白在进化上比较保守,普遍含有一个 Homobox 保守结构域,这对后续的功能验证研究提供可靠的靶向序列。在百脉根 TALE 家族成员中,有 9 条结构域缺失的序列,由于进化过程中基因家族的扩张事件会造成假基因化现象^[22],因此百脉根 TALE 家族可能会存在基因功能的缺失。对百脉根与拟南芥、水稻、大豆 4 种植物的 TALE 转录因子家族成员蛋白序列的进化特点进行分析,发现百脉根 TALE 转录因子较为保守,均为 KNOX 和 BELL 亚族成员; 4 种植物中 BELL 和 KNOX 亚家族的分支非常明显,表明两个亚家族在很早的时候就发生了分化,序列上有较大差异; 百脉根和大豆与水稻分支明显,可能与单、双子叶植物本身的差异有关,相关报道表明复叶植物中 KNOX 基因的表达模式与其他植物中的不一致^[23]。直系同源基因定位发现大豆全基因组复制方式为串联重复^[24],真核生物的串联重复形成基因簇同时拷贝数发生变异^[25],形成大量的旁系同源基因。同时有研究表明,大豆^[26]、百脉根^[27]等豆科植物大约在 5800 万年前发生过一次全基因组复制,但大豆在大约 1300 万年前又发生了一次全基因组复制,这导致 75% 的大豆基因存在多拷贝现象^[26],这也使大豆 TALE 转录因子家族成员在数量上远多于其他植物,而百脉根则与水稻、拟南芥在家族成员数量上较为一致。在染色体分布上,百脉根 TALE 家族成员并不均匀,这可能与响应生物及非生物胁迫的相关基因扩增有密切关系^[28]。百脉根 TALE 家族 40 条蛋白均不含信号肽,表明其没有蛋白质定向转运的功能,而三级结构的差异,也体现了两个亚族的功能差异。通过蛋白质理化性质分析,发现百脉根 TALE 转录因子家族不同成员的蛋白分子质量及氨基酸数量有较大差异,可见 TALE 基因家族对于植物具有复杂的调控方式,不同的分子量、

氨基酸数量、结构域差异可能代表促进或是抑制的调节效果。

百脉根 *TALE* 基因差异表达分析发现 BELL 亚族的 Lj0g3v0034399.1、Lj1g3v4752480.1 主要在芽表达, 而 KONX 亚族的 Lj3g3v1272300.1、Lj6g3v1088990.1 和 BELL 亚族的 Lj0g3v0057839.1 3 条基因主要在根表达, 这可能是由于不同亚族之间主要表达部位存在差异, 这与拟南芥^[29]、水稻^[30] 等多种植物中进行的 *TALE* 基因的组织表达研究结果相似, 可以推测百脉根的 *TALE* 家族与其他植物中的该家族功能相类似。

本研究通过对鉴定出的 40 条百脉根 *TALE* 转录因子家族成员进行生物信息学分析及实验验证, 较为系统地揭示了百脉根 *TALE* 转录因子的理化性质、进化特点和组织表达差异特性, 为今后利用基因工程分析百脉根 *TALE* 转录因子功能奠定基础。

参考文献

- [1] Bellesbois E, Hamant O, Sarah M W, Halima M, Jan T, Ve'ronique P.KNAT6: An arabidopsis homeobox gene involved in meristem activity and organ separation. *Plant Cell*, 2006, 18 (8): 1900-1907
- [2] Aida M, Ishida T, Tasaka M. Shoot apical meristem and cotyledon formation during Arabidopsis embryogenesis: interaction among the *CUP-SHAPED COTYLEDON* and *SHOOT MERISTEMLESS* genes. *Development*, 1999, 126 (8): 1563-1570
- [3] Shani E, Yanai O, Ori N, The role of hormones in shoot apical meristem function. *Current Opinion in Plant Biology*, 2006, 9 (5): 484
- [4] Cnops G, Neyt P, Raes J, Petraruolo M, Nelissen H, Malenica N, Luschnig C, Tietz O, Ditengou F, Palme K, Azmi A, Prinsen E, van Lijsebettens M. The *TORNADO1* and *TORNADO2* genes function in several patterning processes during early leaf development in Arabidopsis thaliana. *Plant Cell*, 2006, 18 (4): 852-866
- [5] Pautot V. Plant development: a *TALE* story. *Comptes rendus-Biologies*, 2010, 333 (4): 371-381
- [6] Smith H M, Hake S, The interaction of two homeobox genes, *BREVIPEDICELLUS* and *PENNYWISE*, regulates internode patterning in the Arabidopsis inflorescence. *Plant Cell*, 2003, 15 (8): 1717-1727
- [7] 郝好, 崔永涛, 钱前, 马伯军, 胡兴明. 水稻同源异形盒基因的研究进展. *中国稻米*, 2016 (1): 1-9
Hao H, Cui Y T, Qian Q, Ma B J, Hu X M. Research progress of homeobox gene in rice. *China Rice*, 2016, 22 (1): 1-9
- [8] 张昉, 赵月明, 丁兵, 齐学军, 解莉. *KNOX1* 基因在植物复叶发育过程中的调控作用. *植物生理学报*, 2015 (11): 1780-1786
Zhang Y, Zhao Y M, Ding B, Qi X J, Xie L N. Regulation of *KNOX1* genes in compound leaf development of plants. *Plant Physiology Journal*, 2015 (11): 1780-1786
- [9] 杨霞, 高金珊, 杨素欣. 豆科复叶发育分子遗传机制的研究进展. *植物生理学报*, 2017 (6): 905-915
Yang X, Gao J S, Yang S X. Progress of molecular mechanism of compound leaf development in legume plants. *Plant Physiology Journal*, 2017 (6): 905-915
- [10] Champagne C E M, Goliber T E, Wojciechowski M F, Mei R W, Townsley B T, Wang K, Paz M M, Geeta R, Sinha N R. Compound leaf development and evolution in the legumes. *Plant Cell*, 2007, 19 (11): 3369-3378
- [11] Choi H K, Mun J H, Kim D J, Zhu H, Baek J M, Mudge J, Roe B, Ellis N, Doyle J, Kiss G B, Young N D, Cook D R. Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101 (43): 15289-15294
- [12] Wang Z, Chen J, Weng L, Li X, Cao X, Hu X, Luo D, Yang J. Multiple components are integrated to determine leaf complexity in *Lotus japonicus*. *Chinese Bulletin of Botany*, 2013, 55 (5): 419
- [13] Jin J, Tian F, Yang D C, Meng Y Q, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Research*, 2017, 45 (Database issue): D1040-D1045
- [14] 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 Research*, 2016, 44 (Web Server issue): W242-W245
- [15] Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones S J, Marra M A. Circos: An information aesthetic for comparative genomics. *Genome Research*, 2009, 19 (9): 1639-1645
- [16] 冯志娟, 徐盛春, 刘娜, 张古文, 胡齐赞, 龚亚明. 植物 TCP 转录因子的作用机理及其应用研究进展. *植物遗传资源学报*, 2018, 19 (1): 112-121
Feng Z J, Xu S C, Liu N, Zhang G W, Hu Q Z, Gong Y M. Molecular Mechanisms and Applications of TCP Transcription Factors in Plants. *Journal of Plant Genetic Resources*, 2018, 19 (1): 112-121
- [17] Chen H, Banerjee A K, Hannapel D J. The tandem complex of BEL and KNOX partners is required for transcriptional repression of *ga20ox1*. *Plant Journal*, 2004, 38 (2): 276-284
- [18] Bhatt A M, Etchells J P, Canales C, Lagodienko A, Dickinson H. VAAMANA-a BEL1-like homeodomain protein, interacts with KNOX proteins BP and STM and regulates inflorescence stem growth in Arabidopsis. *Gene*, 2004, 328 (1): 103-111
- [19] Richardson A E, Rebocho A B, Coen E S. Coen. Ectopic *KNOX* expression affects plant development by altering tissue cell Polarity and Identity. *Plant Cell*, 2016, 28 (9): 2079-2096
- [20] Mahajan A, Bhogale S, Kang I H, Hannapel D J, Banerjee A K. The mRNA of a Knotted1-like transcription factor of potato is phloem mobile. *Plant Molecular Biology*, 2012, 79 (6): 595-608
- [21] Yi P, Soper T J, Woodson S A. RNase Footprinting of Protein Binding Sites on an mRNA Target of Small RNAs//Bacterial Regulatory RNA. New Jersey, USA: Humana Press, 2012: 213
- [22] Magadum S, Banerjee U, Murugan P, Gangapur D, Ravikesavan R. Gene duplication as a major force in evolution.

- Journal of Genetics, 2013, 92(1): 155-161
- [23] Bharathan G, Goliber T E, Moore C, Kessler S, Pham T, Sinha N R. Homologies in leaf form inferred from *KNOX1* gene expression during development. *Science*, 2002, 296(5574): 1858-1860
- [24] Zhao M, Meyers B C, Cai C, Xu W, Ma J. Evolutionary patterns and coevolutionary consequences of miRNA genes and microRNA targets triggered by multiple mechanisms of genomic duplications in soybean. *Plant Cell*, 2015, 27(3): 546
- [25] Lawrence J. Selfish operons: the evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Current Opinion in Genetics & Development*, 1999, 9(6): 642-648
- [26] Schmutz J, Cannon S B, Schlueter J, Ma J, Mitros T, Nelson W, Hyten D L, Song Q, Thelen J J, Cheng J, Xu D, Hellsten U, May G D, Yu Y, Sakurai T, Umezawa T, Bhattacharyya M K, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang X C, Shinozaki K, Nguyen H T, Wing R A, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker R C, Jackson S A. Genome sequence of the palaeopolyploid soybean. *Nature*, 2010, 463(7278): 178-183
- [27] Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S. Genome structure of the legume, *Lotus japonicus*. *Dna Research An International Journal for Rapid Publication of Reports on Genes & Genomes*, 2008, 15(4): 227-239
- [28] Michelmore R W, Meyers B C. Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research*, 1998, 8(11): 1113-1130
- [29] Liberman L M, Sparks E E, Morenorisueno M A, Petricka J J, Benfey P N. MYB36 regulates the transition from proliferation to differentiation in the *Arabidopsis* root. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(39): 12099-12104
- [30] Postmaarsma A D, Verwoert I I, Stronk O P, Koster J, Lamers G E, Hoge J H, Meijer A H. 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*, 1999, 39(2): 257-271