

薄壳山核桃全基因组 LBD 基因家族的生物信息学分析

黄元城, 郭文磊, 王正加

(浙江农林大学 亚热带森林培育国家重点实验室, 浙江 杭州 311300)

摘要: 【目的】研究薄壳山核桃 *Carya illinoensis* LBD 基因家族结构特征、进化模式和在胚发育过程中的表达模式。【方法】运用生物信息学手段鉴定薄壳山核桃 LBD 基因, 分析该基因结构特征、系统发生学关系、显花植物中的进化历史和胚发育过程中 3 个关键阶段的表达模式。【结果】薄壳山核桃全基因组中共鉴定到 52 个候选 LBD 基因。根据基因结构、系统发生学最大似然树和 Motif 分析可分为 3 类: Group I、Group II 和 Group III。多序列比对分析中, 52 个 LBD 基因 LOB 结构域中鉴定出 3 个重要的结构: CX₂CX₆CX₃C 锌指结构、高度保守的甘氨酸 GAS 结构和亮氨酸拉链 (zipper-like) 结构, 并且在 3 类内都分别发生了特异性的变异或者缺失。根据代表性显花植物 LBD 基因家族的系统发生学分析, 从变异程度看 Group I 和 Group II 相对较为保守, 而 Group III 内的所有 LBD 基因共享 1 支较长的分支, 它们已发生了较大的变异, 可能已经分化出新的功能。表达分析结果显示: LBD 基因家族参与调控胚发育过程, 通常控制子叶的发育和形态建成。薄壳山核桃 LBD 基因中又有在整个胚发育过程中都高表达的一簇基因, 这些基因可能在胚发育过程中发挥了更加重要的作用。【结论】薄壳山核桃全基因组中共获得 LBD 基因 52 个, 共可分为 3 个亚家族, 不同的亚家族具有不同的基因结构、蛋白质结构、进化模式和表达模式, 转录组表达分析显示: 不同亚家族之间在胚发育不同阶段具有差异性表达, 它们共同参与调控薄壳山核桃胚发育过程。图 5 表 2 参 47

关键词: 薄壳山核桃; LBD; 转录因子; 系统进化; 基因功能

中图分类号: S722.3; Q75 文献标志码: A 文章编号: 2095-0756(2021)03-0464-12

Genome-wide identification and bioinformatics analysis of LBD family of transcription factors in *Carya illinoensis*

HUANG Yuancheng, GUO Wenlei, WANG Zhengjia

(State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou 311300, Zhejiang, China)

Abstract: [Objective] This study is aimed to investigate the structural features and the evolutionary patterns of the LBD transcription factor family as well as their expression patterns throughout the embryonic development in *Carya illinoensis*. [Method] With bioinformatic software employed, efforts were made to identify the LBD gene and analyze its gene structure characteristics, phylogeny relationships and evolutionary history in flower plants, as well as their expression patterns at the three key stages throughout the embryonic development. [Result] Within the whole genome in *C. illinoensis* there were a total of 52 candidate LBD genes which, in terms of gene structure, maximum likelihood phylogenetic tree and motif analysis, could be classified into 3 categories: Group I, Group II and Group III. After the multiple sequences alignment analysis three important structures in the LOB domains of 52 LBD genes were identified, namely the CX₂CX₆CX₃C zinc finger structure, the highly conserved glycine GAS structure and the leucine zipper-like structure with the occurrence

收稿日期: 2020-07-14; 修回日期: 2020-12-24

基金项目: 浙江省科技厅重大研发专项 (2018C02004); 浙江省农业新品种选育重大科技专项 (2016C02052-13); “十三五”国家重点研发计划项目 (2018YFD1000604)

作者简介: 黄元城 (ORCID: 0000-0002-2388-6057), 从事经济林培育研究。E-mail: 834194037@qq.com。通信作者: 王正加 (ORCID: 0000-0002-6511-6771), 教授, 博士, 博士生导师, 从事经济林遗传育种和植物发育生物学研究。E-mail: wzjh21@163.com

of specific mutations or deletions in each of them. According to the LBD gene family phylogenetic analysis of representative flowering plants, Group I and Group II are relatively conservative, while all LBD genes in Group III share a longer branch which indicates that they have undergone greater variation and new functions have been differentiated. The expression analysis showed that the LBD gene family participates in the function of regulating the development of embryos, usually with control over the development and morphogenesis of cotyledons and there is a cluster of genes in the LBD gene of hickory pecans that are highly expressed during the entire embryo development process playing a highly significant role. [Conclusion] A total of 52 LBD genes were obtained in the whole genome of *C. illinoensis*, which can be divided into three different subfamilies with different gene structures, protein structures, evolutionary patterns, and expression patterns, and these subfamilies, according to the transcriptome expression analysis, are differentially expressed at different stages of embryonic development, and involved in the regulation of *C. illinoensis* embryo development collectively. [Ch, 5 fig. 2 tab. 47 ref.]

Key words: *Carya illinoensis*; LBD; transcription factor; phylogenetics; gene function

LBD 转录因子是一类在其蛋白质 N 端具有侧生器官边界 (LOB) 蛋白质结构域的植物特有转录因子^[1]。SHUAI 等^[2]首次发现了 LBD 基因,发现它在植物侧生器官的边界细胞中表达并且参与侧生器官的发育形成。根据侧生器官边界 (LOB) 域的结构,拟南芥 *Arabidopsis thaliana* 中的 LBD 转录因子家族通常可被分为 2 类。类 1(Group 1) 具有 1 个完整的高度保守的 CX₂CX₆CX₃C 锌指结构,通常能够结合特定 DNA 序列。同时类 1 的 LBD 基因具有高度保守的甘氨酸 GAS 结构和亮氨酸拉链 (zipper-like) 结构 (LX₆LX₃LX₆L), 它们之间会形成卷曲-螺旋结构的蛋白质并相互作用。而类 2(Group 2) 只含有 1 个保守的 CX₂CX₆CX₃C 锌指结构^[1-4]。大多数的 LBD 转录因子基因从属于类 1。类 2 的 LBD 蛋白质往往保存有 1 个可能不具有功能的残缺的亮氨酸拉链结构,它会形成 1 个盘绕-线圈结构^[4]。最近,CHEN 等^[5]通过研究小麦 *Triticum aestivum* LBD 蛋白质中 LOB 结构域的蛋白质结晶学,揭示了之前提到的 CX₂CX₆CX₃C 锌指结构实际是 C₄ 型锌指结构。C₄ 锌指结构常常与 GAS 结构以及 α₄ 和 α₅ 之间的垂直构象共同作用,从而精确识别 DNA,并能对结合位点的空间构型进行定位。最初的一些 LBD 基因功能研究表明:LBD 基因通常在侧生器官底部的边界新生细胞中表达,在器官分化和侧生器官发育中具有潜在功能^[1]。同时,LBD 基因还能够参与植物花青素、氮元素代谢和器官再生等关键过程^[6]。类 1 的 LBD 基因大多数参与植物发育过程^[4,7]和由生长素信号转导级联反应介导的侧根形成过程。相反,类 2 基因往往作为花青素合成的阻遏物和有效性信号参与了新陈代谢过程^[8]。在拟南芥 LBD 转录因子的表达模式研究中,病原体几乎诱导了类 2 的所有 LBD 基因的表达,这表明 LBD 基因能够在植物病原防御反应中发挥作用^[9]。LBD 基因参与到拟南芥许多组织发育过程中,比如叶片发育^[10]、根部侧生器官发育^[11-13]、细胞分裂素信号转导^[14]和赤霉素途径^[15]。值得一提的是,拟南芥中的 *AthLBD16* 基因会促进侧根的起始发育^[16],*AthLBD29* 则能够抑制拟南芥茎纤维壁增厚的生长素信号转导过程^[17]。在尖孢镰刀菌 *Fusarium oxysporum* 中 *AtLBD20* 作为易感基因似乎能够调节茉莉酸 (JA) 信号传导和激活尖孢镰刀菌中 JA 信号转导过程依赖的反应^[18]。除拟南芥之外 LBD 基因功能在许多物种中同样被研究过,如 *OsIG1* 基因会参与配子发生过程并影响水稻 *Oryza sativa* 的花器官数量^[19]; *MdLBD13* 蛋白质则可以抑制苹果 *Malus × domestica* 中的花色苷合成和氮吸收^[20]。与此同时,在许多物种中也开展了利用基因组数据资源对 LBD 基因家族鉴定和分析的研究,如茶树 *Camellia sinensis*^[21]、马铃薯 *Solanum tuberosum*^[22]、桉树 *Eucalyptus robusta*^[23]、芸薹 *Brassica campestris*^[24]、葡萄 *Vitis vinifera*^[25]、大豆 *Glycine max*^[26]。但是目前薄壳山核桃 *Carya illinoensis* 中的 LBD 基因研究却鲜有报道。在薄壳山核桃全基因组测序组装完成后 LBD 基因家族仍然没有进行系统的研究^[27]。本研究通过生物信息学手段鉴定了薄壳山核桃全基因组内的 LBD 基因,分析其基因结构、基序 (motif) 分布、转录组表达模式并构建系统进化树,预测 LBD 基因家族在薄壳山核桃中可能的功能,并为进一步研究提供理论基础。

1 材料与方法

1.1 多物种全基因组蛋白质序列获取

薄壳山核桃、山核桃 *Carya cathayensis*、核桃 *Juglans regia* 的全基因组蛋白质序列来源于胡桃科 Juglandaceae 数据库 (<http://www.juglandaceae.net/>)^[28]。银杏 *Ginkgo biloba*、无油樟 *Amborella trichopoda*、蓝星睡莲 *Nymphaea colorata*、大豆、葡萄、水稻的全基因组蛋白质序列来源于美国国立生物技术信息中心 (NCBI)。拟南芥的全基因组蛋白质序列来源于拟南芥信息资源库 (TAIR, <https://www.arabidopsis.org/>)。

1.2 LBD 转录因子基因的鉴定与蛋白质理化性质分析

以 LBD 为关键词, 搜索获取 LBD 在拟南芥中的 10 条核酸编码序列 (CDS), 并下载这 10 条基因的 FASTA 格式的核酸序列。将获得的 10 条拟南芥编码序列通过本地 NCBI-BLAST 2.9.0 软件的 BLASTX 程序与 1.1 中获得的全基因组蛋白质序列进行同源比对, 在 Pfam 数据库^[29]中下载 LOB 结构域的种子文件, 得到的候选 LBD 蛋白质序列用 HMMER 软件^[30]使用隐马尔科夫模型算法筛选出 LBD 基因, 利用 ExPASy 蛋白质在线分析网站对蛋白质序列进行理化性质的鉴定和分析。

1.3 LBD 基因组共线性区块分析

通过 MCSCAN 软件^[31], 将薄壳山核桃全基因组核酸编码序列 CDS 文件和基因组注释 GFF 文件输入执行 Pairwise Synteny Search 程序, 得到的 Anchor Pairwise 结果以 30 个基因为阈值过滤小片段 block。在最后的 gene block 中寻找 1.2 中鉴定到的 LBD 基因。

1.4 LBD 基因多重序列比对

利用 ENSEMBL 的 CLUSTALW 软件^[32]对获取的 52 条薄壳山核桃 LBD 基因进行多重序列比对, 将比对后的 ALN 文件输入至 ENDscript 网站进行多序列比对的可视化。

1.5 LBD 基因系统发育、基因结构和基序分析

利用 MUSCLE v3.8.31 软件^[33]分别对获取的 52 条薄壳山核桃和 547 条多物种 LBD 基因的蛋白质序列进行多序列比对。获得的多序列比对结果文件使用 FastTree V2.1.11 软件^[34]的 JTT+CAT 算法^[35]构建最大似然树, 获得的树文件通过 Evolview 在线网站^[36]可视化。在胡桃科数据库下载薄壳山核桃基因组结构注释 GFF 文件, 并利用 Gene Structure Display Server 2.0 在线网站^[37]可视化 LBD 基因结构图。在 MEME 网站上传 52 条薄壳山核桃蛋白质序列分析基序。

1.6 LBD 基因表达模式分析

转录组原始测序数据从 NCBI (<https://www.ncbi.nlm.nih.gov/>) PRJNA435846 下的 SRA 数据库获取: SRR6793964、SRR6793963、SRR6793962、SRR6793961、SRR6793960、SRR6793959、SRR6793958、SRR6793957、SRR6793956、SRR6793955。从 GIGADB (<http://gigadb.org/>) 获取薄壳山核桃全基因组序列和基因组注释 GFF 文件, 利用 fastp 软件进行质控和过滤, STAR 软件进行序列比对, RSEM 软件对基因进行定量分析, 计算 TPM 值并生成基因表达数据框。得到的表达数据通过 R 软件将表达值进行对数转换后利用 pheatmap 包绘制表达热图。

2 结果与分析

2.1 LBD 基因鉴定与蛋白质理化性质分析

通过 BLASTX 初筛出 LBD 的同源基因, 经 HMM-SEARCH 程序^[29]分析和筛选, 供试的 10 个物种中共获得 551 条 LBD 基因 (表 1)。其中大豆中 LBD 基因的数量是无油樟的 4.46 倍、核桃的 1.61 倍, 推测这可能和大豆是由古四倍体演化而来的二倍体有关^[38]。根据薄壳山核桃基因组 GFF 注释文件, 发现有 *pecanLBD50*、*pecanLBD51* 和 *pecanLBD52* 等 3 个基因是位于 scaffold123681 的短串联重复基因。

MCSCAN 分析显示: 薄壳山核桃基因组共有 211 个共线性的区块。在这些共线性的区块中找到了 13 对在可能全基因组加倍事件中发生重复的同源 LBD 基因对: *pecanLBD1* 和 *pecanLBD25*、*pecanLBD4* 和 *pecanLBD17*、*pecanLBD5* 和 *pecanLBD16*、*pecanLBD11* 和 *pecanLBD33*、*pecanLBD11* 和 *pecanLBD32*、*pecanLBD12* 和 *pecanLBD33*、*pecanLBD12* 和 *pecanLBD32*、*pecanLBD14* 和 *pecanLBD43*、*pecanLBD14* 和 *pecanLBD42*、*pecanLBD19* 和 *pecanLBD26*、*pecanLBD20* 和 *pecanLBD29*、*pecanLBD21*

表 1 多物种 LBD 基因鉴定结果

Table 1 Identify result of LBD genes in multi-species

鉴定程序	基因数量/个									
	银杏	无油樟	蓝星睡莲	水稻	拟南芥	葡萄	大豆	核桃	山核桃	薄壳山核桃
BLASTX	90	45	51	61	91	98	153	87	89	80
HMMER	50	26	35	29	58	57	116	72	56	52

和 *pecanLBD36*、*pecanLBD34* 和 *pecanLBD41*。共线性分析共得到基因 23 个，占有所有 LBD 基因的 44.2%，这说明了全基因组加倍事件使 LBD 基因家族得到了扩张。

通过理化性质分析 (表 2)：在氨基酸数量、分子量、等电点、不稳定系数和脂肪系数等方面存在差异。氨基酸数量为 92~326 个，大部分为 170~300 个；等电点为 4.48~9.85，>7.5 的有 24 个，多在碱性范围内；不稳定系数<40 的有 51 个，为稳定蛋白质；>40 仅 1 个，为不稳定蛋白质；蛋白质的脂肪系数大多<100，为疏水性蛋白质，仅有 *pecanLBD50* 脂肪系数>100，为亲水性蛋白质。

2.2 薄壳山核桃 LBD 基因系统发育、基因结构和基序分析

如图 1 所示：52 个薄壳山核桃 LBD 基因聚成 3 类：Group I、Group II、Group III。其中，Group I 含有最多的 33 个 LBD 基因，它们可能承担了 LBD 基因促进侧生器官发育的主要功能。Group I 中有 1 对蛋白质序列完全相同的基因 *pecanLBD22* 和 *pecanLBD48*，导致它们在进化树上没有显示变异，而 *pecanLBD41* 的进化枝长度达到 3.7346，是 Group I 中相对于其他 LBD 基因变异程度最大的。Group II 细分为 3 支基因簇，其中 Sub group I 和 Sub group III 的进化枝相对较短，因此序列和功能变异不大；而 Sub group II 有明显的进化变异，其中 *pecanLBD6* 的枝长为 6.2748，是 Sub group II 中变异最大的基因，它可能已经出现了功能上的分化变异。Group III 是 3 类中最保守的一支，其中的 *pecanLBD9* 基因的枝长在所有 52 个 LBD 基因中最长，达 6.764 2。

如图 2 中 52 个薄壳山核桃 LBD 基因基序分析结果所示：共发现了 10 个基序并命名为 Motif1~Motif10。对 LBD 转录因子家族较重要的有 3 个基序：Motif1 含 GAS(Gly-Ala-Ser) 甘氨酸保守结构；Motif 2 拥有完整的高度保守的 CX₂CX₆CX₃C 锌指结构，通常具有结合特定 DNA 序列的能力；Motif 3 为亮氨酸拉链结构 (LX₆LX₃LX₆L)。Motif1 和 Motif3 往往会形成卷曲-螺旋结构的蛋白质并相互作用。52 个薄壳山核桃 LBD 基因中的 39 个具有完整的 Motif1、Motif2 和 Motif3 的顺序基序结构；而 *pecanLBD34*、*pecanLBD41* 和 *pecanLBD9* 在进化过程中丢失了 Motif1 的 GAS 甘氨酸保守结构；*pecanLBD6*、*pecanLBD19*、*pecanLBD26*、*pecanLBD45*、*pecanLBD37*、*pecanLBD2* 和 *pecanLBD3* 丢失了 Motif3 亮氨酸 zipper-like 结构 (LX₆LX₃LX₆L)，尽管它们仍具有 GAS 结构但已经丧失了和 Motif1 结合生成卷曲-螺旋蛋白质结构的功能；*pecanLBD50* 则丢失了所有的基序，可能丧失了 LBD 基因的基本功能。

基因结构分析显示：薄壳山核桃 LBD 基因一般具有 1~3 个外显子。其中 *pecanLBD41*~*pecanLBD52* 都只具有 1 个外显子 (除 *pecanLBD47* 有 2 个外显子除外)，并且都属于 Group I；具有 3 个外显子的基因只有 *pecanLBD3*、*pecanLBD34* 和 *pecanLBD9*，在每类都有部分；剩余的 LBD 基因都为 2 个外显子，占有所有薄壳山核桃 LBD 基因的 67.3%。因此，薄壳山核桃 LBD 转录因子家族的大多数基因较为保守，有着相似的基序、基因结构，执行并发挥 LBD 的主要功能，而其发生变异的小部分可能已经丢失了原来的基因功能或者特化出新的功能。

2.3 薄壳山核桃 LOB 结构域多序列比对

对 52 个薄壳山核桃 LBD 转录因子基因的 LOB 蛋白质结构域多序列比对 (图 3) 发现：LOB 蛋白质结构域主要由 3 个保守序列模式组成：长度为 16 个氨基酸的 CX₂CX₆CX₃C 锌指结构、长度为 50 个氨基酸的甘氨酸 GAS 结构和 LX₆LX₃LX₆L 亮氨酸拉链结构。其中 *pecanLBD50* 和 *pecanLBD21* 完全缺失了锌指结构和部分的甘氨酸 GAS 结构，而其他的 LBD 的锌指结构则全部保存下来没有发生缺失或者突变，较为保守。甘氨酸 GAS 结构相对锌指结构变异程度较大，Group I 中 *pecanLBD23*~*pecanLBD47*、*pecanLBD10*、*pecanLBD51* 发生了 1~2 个氨基酸的变异，Group II 中 *pecanLBD6*、*pecanLBD19*、*pecanLBD26*、*pecanLBD45*、*pecanLBD37*、*pecanLBD2*、*pecanLBD3* 等 7 个转录因子 GAS 中的丙氨酸突

表2 薄壳山核桃 LBD 转录因子家族蛋白质理化性质

Table 2 Physicochemical properties of LBD transcription factor family protein in *C. illinoensis*

蛋白质	氨基酸残基数/个	分子量/kD	理论等电点	酸性氨基酸/个	碱性氨基酸/个	不稳定系数	脂肪系数	总平均亲水性
pecanLBD1	210	23 309.19	8.24	17	19	47.65	73.52	-0.485
pecanLBD2	208	22 353.77	7.55	20	21	58.32	88.65	0
pecanLBD3	288	31 617.01	8.99	29	35	61.92	76.88	-0.326
pecanLBD4	223	24 069.37	7.69	17	18	73.89	77.89	-0.171
pecanLBD5	188	20 808.56	7.63	17	18	67.24	69.15	-0.490
pecanLBD6	260	28 451.22	8.23	31	34	52.55	69.38	-0.501
pecanLBD7	156	17 540.20	9.03	12	18	55.87	78.91	-0.235
pecanLBD8	172	18 611.24	8.59	14	17	61.24	70.35	-0.238
pecanLBD9	276	30 985.11	8.55	28	32	30.84	76.96	-0.522
pecanLBD10	326	36 029.86	6.91	23	22	64.45	77.58	-0.472
pecanLBD11	204	22 376.57	7.53	18	19	60.78	77.99	-0.233
pecanLBD12	232	25 857.68	5.90	21	14	50.52	65.22	-0.425
pecanLBD13	170	18 808.03	6.42	18	17	65.04	57.47	-0.607
pecanLBD14	202	21 823.68	8.59	14	17	57.41	80.20	-0.175
pecanLBD15	203	23 706.08	8.15	27	29	52.00	77.39	-0.556
pecanLBD16	191	20 998.77	6.08	19	17	70.05	72.04	-0.423
pecanLBD17	230	25 054.58	9.22	17	23	78.94	70.04	-0.284
pecanLBD18	172	18 626.23	8.82	14	18	67.64	73.14	-0.296
pecanLBD19	306	33 113.72	7.95	33	35	46.03	81.54	-0.280
pecanLBD20	162	18 029.72	6.70	14	14	51.76	88.52	-0.133
pecanLBD21	127	14 182.02	4.48	14	7	36.18	92.20	-0.246
pecanLBD22	176	19 971.98	5.17	25	18	44.88	80.34	-0.186
pecanLBD23	130	15 042.74	9.84	10	25	45.43	77.23	-0.541
pecanLBD24	162	18 285.75	8.23	13	15	57.65	78.40	-0.337
pecanLBD25	214	23 771.70	8.25	17	19	49.42	74.44	-0.472
pecanLBD26	310	33 384.03	8.22	34	37	47.44	80.87	-0.302
pecanLBD27	168	18 892.28	6.28	17	13	72.86	63.87	-0.476
pecanLBD28	215	23 539.68	5.33	18	14	75.47	70.88	-0.272
pecanLBD29	167	18 504.18	6.94	15	15	59.00	81.80	-0.217
pecanLBD30	176	19 230.73	6.50	18	17	69.15	73.75	-0.242
pecanLBD31	169	18 881.58	6.49	17	16	51.28	79.59	-0.219
pecanLBD32	236	26 157.91	5.74	22	13	46.73	65.76	-0.405
pecanLBD33	213	23 405.50	7.06	19	19	65.23	67.37	-0.380
pecanLBD34	326	37 266.20	5.47	54	35	49.05	74.29	-0.813
pecanLBD35	251	27 236.49	6.44	24	22	59.55	80.08	-0.322
pecanLBD36	228	24 949.42	6.29	17	16	83.40	71.05	-0.271
pecanLBD37	228	24 709.22	8.06	22	24	59.20	76.14	-0.244
pecanLBD38	216	23 449.53	6.03	17	13	64.11	75.05	-0.227
pecanLBD39	158	17 663.15	6.27	15	13	53.44	80.25	-0.243
pecanLBD40	221	24 171.04	4.97	21	12	58.45	78.60	-0.143
pecanLBD41	265	30 244.09	6.42	33	29	46.86	77.66	-0.594
pecanLBD42	237	24 801.38	8.20	13	15	69.07	79.87	-0.065
pecanLBD43	213	23 452.55	6.74	19	18	56.45	80.61	-0.203
pecanLBD44	224	24 373.77	8.93	17	22	69.41	73.62	-0.276
pecanLBD45	202	21 927.15	8.56	19	23	54.66	83.47	-0.159
pecanLBD46	221	24 114.37	9.01	18	23	73.18	66.29	-0.421
pecanLBD47	289	32 466.21	5.27	29	21	67.19	65.92	-0.600
pecanLBD48	176	19 971.98	5.17	25	18	44.88	80.34	-0.186

表 2 (续)

蛋白质	氨基酸残基数/个	分子量/kD	理论等电点	酸性氨基酸/个	碱性氨基酸/个	不稳定系数	脂肪系数	总平均亲水性
pecanLBD49	170	18 770.95	6.19	19	17	60.44	55.18	-0.615
pecanLBD50	92	10 319.08	6.26	11	10	39.17	115.65	0.018
pecanLBD51	313	34 650.03	7.29	20	20	64.34	73.07	-0.490
pecanLBD52	320	35 091.25	6.83	23	21	66.59	67.53	-0.593

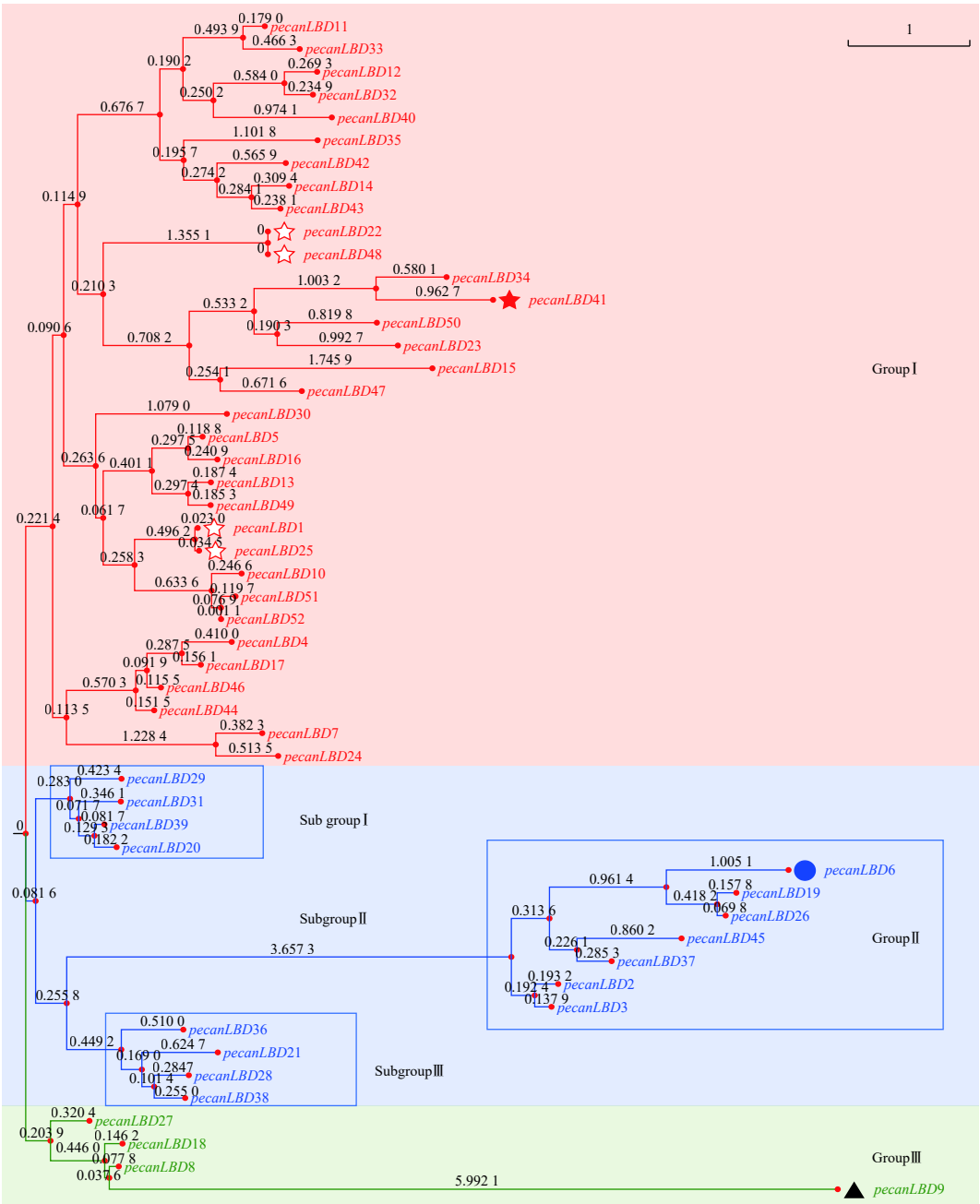


图 1 薄壳山核桃 LBD 基因家族系统进化树

Figure 1 Phylogenetic tree constructed based on the full-length sequences of pecan LBD genes using JTT+CAT algorithm with FastTree software

变为精氨酸、丝氨酸突变为丙氨酸，52 个 LBD 转录因子中有 12 个发生了突变，2 个缺失了 GAS 结构。LX₆LX₃LX₆L 亮氨酸拉链结构中，上述 7 个转录因子的第 1 个亮氨酸位置都突变为甘氨酸，第 3 个亮氨酸位置突变为甘氨酸或苏氨酸，有 5 个转录因子的第 2 个亮氨酸位置发生了突变，由此可见上述 7 个 LBD 转录因子与进化模式 (图 1) 一致，已经发生了较大程度的变异，可能已经分化出新的功能。



图2 薄壳山核桃 LBD 转录因子家族系统发育树和保守蛋白质基序结构

Figure 2 Phylogenetic tree constructed based on the full-length sequences of pecan LBD genes using JTT+CAT algorithm with FastTree software



图3 薄壳山核桃 LBD 转录因子家族 LOB 蛋白质结构域多序列比对

Figure 3 Multiple sequence alignment of LOB domain in *C. illinoensis* transcription factor family

pecanLBD15、pecanLBD47、pecanLBD10、pecanLBD51、pecanLBD52、pecanLBD7 和 pecanLBD24 的第 2 个亮氨酸位置发生了突变，第 3 个亮氨酸位置只有 pecanLBD40、pecanLBD23 和 pecanLBD10 发生了突变，相对保守。

2.4 显花植物 LBD 转录因子家族系统发育分析

由图 4 可见：与薄壳山核桃 LBD 系统发育树相似，图 4 中各物种 LBD 转录因子可以聚为 3 类：Group I、Group II 和 Group III。其中 Group I 有 314 个 LBD 基因，为总数的 57.0%，占比最多；Group II 有 166 个 LBD 基因，占总数的 30.1%；而 Group III 的基因数量最少，只有 71 个，占总数的 12.9%。从表 1 鉴定到的 LBD 转录因子的数量来看，裸子植物银杏的数量是早期被子植物类群的 2 倍，这与银杏最近一次特异性的 WGD 事件有关^[39]；早期被子植物类群和单子叶植物水稻的 LBD 基因数量接近；而双子叶植物的 LBD 基因数量明显又发生了 1 次倍增。从分布来看，Group I、Group II 和 Group III 中每个植物类群都存在 LBD 基因，因此 LBD 转录因子家族可能在裸子植物和被子植物未分化前已经分为 3 类，分化后 3 类 LBD 继续各自进化。从变异程度看 Group I 和 Group II 相对较为保守，而 Group III 内的所有 LBD 基因共享一支较长的分支，说明它们已发生了较大的变异，可能已经分化出新的功能。

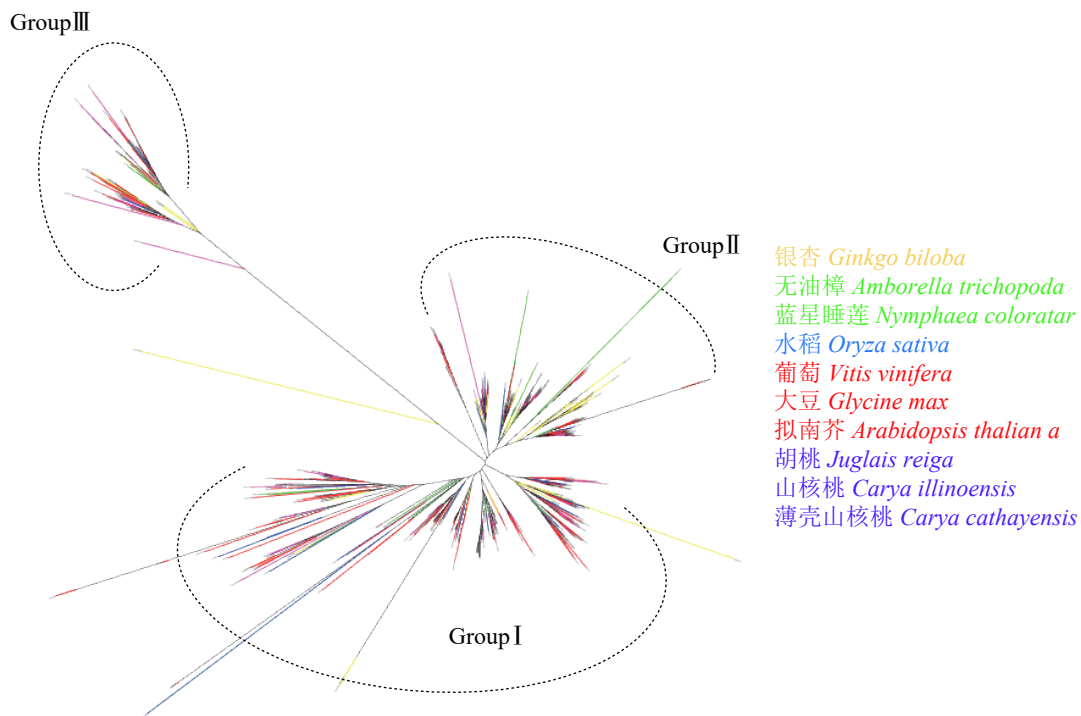


图 4 利用从裸子植物银杏到高等植物 551 个 LBD 基因全长蛋白质序列构建系统发育树

Figure 4 Phylogenetic tree constructed based on the full-length sequences of 551 LBD genes from gymnosperm ginkgo to higher angiosperms

2.5 薄壳山核桃 LBD 基因组织表达模式分析

为了探究 LBD 转录因子家族在胚发育的关键过程中的作用，从 NCBI 网站的 SRA 原始测序数据库中下载了薄壳山核桃胚成熟过程中 3 个关键时期的原始测序数据：子叶伸展早期 (early extended stage of cotyledon development)、子叶完全伸展期 (fully extended stage of cotyledon development)、胚完全成熟期 (fully matured stage of embryo)，并进行分析和绘制聚类热图 (图 5)。不同 LBD 基因和不同发育阶段之间的表达模式有显著差异。基因表达量可以聚为 4 类：① pecanLBD2、pecanLBD19 和 pecanLBD26 在 3 个阶段的表达量比较高；② pecanLBD30、pecanLBD3 和 pecanLBD37 在子叶伸展早期和子叶完全伸展期表达；③ pecanLBD11、pecanLBD18、pecanLBD8、pecanLBD25、pecanLBD6、pecanLBD42、pecanLBD13、pecanLBD52、pecanLBD17、pecanLBD35、pecanLBD1、pecanLBD43、pecanLBD29 和 pecanLBD44 仅在子叶伸展早期有表达；④ 其余的 LBD 基因在 3 个阶段都不表达。从 2.2 中系统发育分析的结果来看：③ 类表达的 14 个 LBD 基因 (78.6%)，属于较为保守的 Group I，因此它们在胚发育过程中主要参与经典的器官分化过程，加快子叶边界新生的细胞分化。而① 类表达的基因都是进化枝较长、变异程度较大的

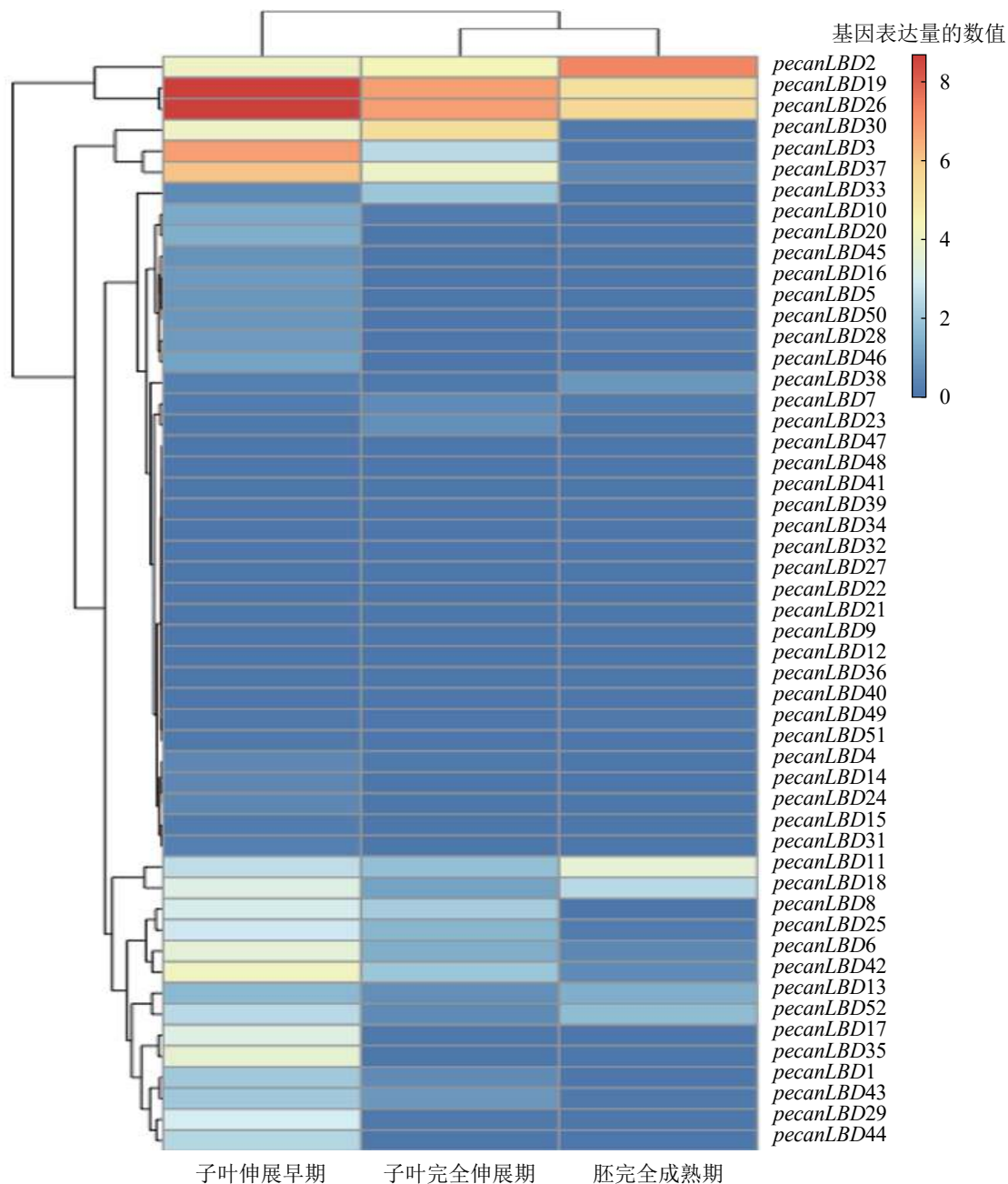


图5 薄壳山核桃 LBD 转录因子家族胚发育表达模式

Figure 5 Transcriptome expression profile of LBD gene family in embryo development

Group II 中 Sub Group II 的成员, 可见 Group II 中 Sub Group II 的 LBD 成员已经发生了较大程度的变异, 它们参与到胚发育的全程, 可能与胚中营养物质的积累过程和胚组织分生发育的过程密切相关。②类中 *pecanLBD30* 和 *pecanLBD3* 属于 Group II, *pecanLBD37* 属于 Group I, 它们在功能上可能介于①类和③类, 即既参与了器官分化、细胞分裂的过程调控, 又和胚中营养物质积累的过程相关。总体上, 薄壳山核桃胚发育过程中 LBD 基因功能较为保守, 主要参与新生细胞分裂分化、子叶形态建成等, 但 Group II LBD 基因的变异程度较大, 可能进化出新的功能, 在胚发育过程中发挥更加关键的作用。

3 结论与讨论

本研究分别在银杏、无油樟、蓝星睡莲、水稻、拟南芥、葡萄、大豆、核桃、山核桃和薄壳山核桃全基因组蛋白质序列中系统鉴定了 551 个 LBD 基因。这些 LBD 基因包含典型的 LOB 结构域, 即 1 个含有 DNA 结合活性所需的 (CX₂CX₆CX₃C) 锌指结构、1 个 Gly-Ala-Ser(GAS) 结构和 1 个负责蛋白质二聚体化的亮氨酸拉链 (LX₆LX₃LX₆L) 结构。根据 LOB 结构域的组成模式和系统发育学分析结果。薄壳山核

桃 52 个 LBD 基因可分为 Group I、Group II 和 Group III 3 类。33 个 LBD 基因被划分为 Group I, 15 个被划分为 Group II, 4 个为 Group III。根据进化枝的长度, Group I 是数量较多同时也比较保守的一支, 而 Group II 和 Group III 都发生了一定程度的变异。

新基因功能产生或分化的一大动力是基因重复, 基因重复对物种提高环境适应性至关重要^[40]。在进化过程中, 重复的基因可能会经历功能分化、亚功能分化或功能丢失等多种过程^[41]。基因重复通常会导致基因家族扩张^[42]。为了揭示薄壳山核桃 LBD 基因家族的复制机制, 利用 MCScanX 对薄壳山核桃的基因组共线性区块内成对的 LBD 基因进行筛选。根据目前的基因组内 LBD 基因位置分析结果, 薄壳山核桃基因组内只存在 1 个连续 3 个 LBD 基因的串联重复事件, 但是存在 13 对 23 个由全基因组加倍产生的 LBD 基因共线性基因对。这说明全基因组加倍在 LBD 基因家族进化中起到了主导性的推动作用。

本研究构建了以裸子植物(银杏)、早期被子植物(无油樟、睡莲)、单子叶植物(水稻)、双子叶植物(拟南芥、葡萄、大豆)、近缘种植物(核桃、山核桃)和薄壳山核桃为显花植物中主要类群的 LBD 基因系统发育树。显花植物 LBD 系统发育树与薄壳山核桃 LBD 相似, 同样可以聚为 3 类。从分布来看, Group I、Group II 和 Group III 中每个植物都存在 LBD 基因, 因此 LBD 转录因子家族可能在裸子植物和被子植物未分化前已经分化为 3 类。从变异程度看, Group I 和 Group II 相对较为保守, 而 Group III 内的所有 LBD 基因共享一支较长的分支, 说明已发生了较大的变异, 可能已经分化出新的功能。

LBD 蛋白质协调了很多植物发育过程, 并对环境刺激作出反应, 特别是在调控侧根器官发育和代谢过程中起着至关重要的作用。例如 LBD18 通过抑制 LBD18 DNA-结合活性控制侧根发育^[43]。一般来说, 主根受到侧根过度生长的影响, 会降低植物提取水分和养分的能力。在土壤中蔓延的侧根过多, 将增加被病原体攻击的可能性。以前的一些工作已经证明了这些预测。如 LBD1 通过抑制主根生长和维持侧根萌发来控制盐胁迫下的根结构^[44]。LBD 基因由于其对一系列下游基因的转录调控, 成为植物病原体入侵的关键分子靶基因^[7]。

在薄壳山核桃研究中胚的发育是非常重要的阶段^[45], 特别是胚内蛋白质、不饱和脂肪酸、淀粉等重要营养物质累积过程是薄壳山核桃研究的热点^[46-47]。本研究分析 LBD 转录因子家族在薄壳山核桃子叶伸展早期、子叶完全伸展期和胚完全成熟期 3 个时期的转录数据发现: 薄壳山核桃 LBD 基因存在 4 种表达模式。可见 LBD 基因具有参与调控胚发育过程的功能, 通常基于在侧生器官底部的边界新生的细胞中调控细胞分裂分化来控制子叶的发育和形态建成。薄壳山核桃 LBD 基因中有在整个胚发育过程中都高表达的一簇基因, 这些基因在胚发育过程中发挥了更加重要的作用, 因此有必要对这一簇 LBD 基因进行更为深入的基因功能研究, 可为进一步改良薄壳山核桃遗传性状提供基础。

4 参考文献

- [1] IWAKAWA H, UENO Y, SEMIARTI E, *et al.* The *ASYMMETRIC LEAVES2* gene of *Arabidopsis thaliana*, required for formation of a symmetric flat leaf lamina, encodes a member of a novel family of proteins characterized by cysteine repeats and a leucine zipper [J]. *Plant Cell Physiol*, 2002, **43**(5): 467–478.
- [2] SHUAI Bin, REYNAGA-PÉÑA C G, SPRINGER P S. The *Lateral organ boundaries* gene defines a novel, plant-specific gene family [J]. *Plant Physiol*, 2002, **129**(2): 747–761.
- [3] LANDSCHULZ W H, JOHNSON P F, MCKNIGHT S L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins [J]. *Science*, 1988, **240**(4860): 1759–1764.
- [4] MAJER C, HOCHHOLDINGER F. Defining the boundaries: structure and function of LOB domain proteins [J]. *Trends Plant Sci*, 2011, **16**(1): 47–52.
- [5] CHEN Weifei, WEI Xiaobin, RETY S, *et al.* Structural analysis reveals a “molecular calipers” mechanism for a lateral organ boundaries domain transcription factor protein from wheat [J]. *J Biol Chem*, 2019, **294**(1): 142–156.
- [6] THATCHER L F, POWELL J J, AITKEN E A B, *et al.* The lateral organ boundaries domain transcription factor LBD20 functions in fusarium wilt susceptibility and jasmonate signaling in *Arabidopsis* [J]. *Plant Physiol*, 2012, **160**(1): 407–418.
- [7] XU Changzheng, LUO Feng, HOCHHOLDINGER F. LOB domain proteins: beyond lateral organ boundaries [J]. *Trends Plant Sci*, 2016, **21**(2): 159–167.

- [8] RUBIN G, TOHGE T, MATSUDA F, *et al.* Members of the *LBD* family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in *Arabidopsis* [J]. *Plant Cell*, 2009, **21**(11): 3567 – 3584.
- [9] THATCHER L F, KAZAN K, MANNERS J M. Lateral organ boundaries domain transcription factors: new roles in plant defense [J]. *Plant Signal Behav*, 2012, **7**(12): 1702 – 1704.
- [10] SEMIARTI E, UENO Y, TSUKAYA H, *et al.* The *ASYMMETRIC LEAVES2* gene of *Arabidopsis thaliana* regulates formation of a symmetric lamina, establishment of venation and repression of meristem-related homeobox genes in leaves [J]. *Development*, 2001, **128**(10): 1771 – 1783.
- [11] GOH T, TOYOKURA K, YAMAGUCHI N, *et al.* Lateral root initiation requires the sequential induction of transcription factors LBD16 and PUCHI in *Arabidopsis thaliana* [J]. *New Phytol*, 2019, **224**(2): 749 – 760.
- [12] CHO C, JEON E, PANDEY S K, *et al.* LBD13 positively regulates lateral root formation in *Arabidopsis* [J]. *Planta*, 2019, **249**(4): 1251 – 1258.
- [13] LEE H W, CHO C, PANDEY S K, *et al.* LBD16 and LBD18 acting downstream of *ARF7* and *ARF19* are involved in adventitious root formation in *Arabidopsis* [J]. *BMC Plant Biol*, 2019, **19**: 46. doi: [10.1186/s12870-019-1659-4](https://doi.org/10.1186/s12870-019-1659-4).
- [14] NAITO T, YAMASHINO T, KIBA T, *et al.* A link between cytokinin and *ASL9* (*asymmetric leaves 2 Like 9*) that belongs to the *AS2/LOB* (*lateral organ boundaries*) family genes in *Arabidopsis thaliana* [J]. *Biosci Biotechnol Biochem*, 2007, **71**(5): 1269 – 1278.
- [15] ZENTELLA R, ZHANG Zhonglin, PARK M, *et al.* Global analysis of DELLA direct targets in early gibberellin signaling in *Arabidopsis* [J]. *Plant Cell*, 2007, **19**(10): 3037 – 3057.
- [16] GOH T, KASAHARA H, MIMURA T, *et al.* Multiple AUX/IAA-ARF modules regulate lateral root formation: the role of *Arabidopsis* *SHY2/IAA3*-mediated auxin signalling [J]. *Phil Trans Roy Soc B Biol Sci*, 2012, **367**(1595): 1461 – 1468.
- [17] LEE K H, DU Qian, ZHUO Chunliu, *et al.* LBD29-involved auxin signaling represses NAC master regulators and fiber wall biosynthesis [J]. *Plant Physiol*, 2019, **181**(2): 595 – 608.
- [18] FENG Zhenhua, ZHU Jian, DU Xiling, *et al.* Effects of three auxin-inducible LBD members on lateral root formation in *Arabidopsis thaliana* [J]. *Planta*, 2012, **236**(4): 1227 – 1237.
- [19] ZHANG Jingrong, TANG Wei, HUANG Yulan, *et al.* Down-regulation of a *LBD*-like gene, *OsIG1*, leads to occurrence of unusual double ovules and developmental abnormalities of various floral organs and megagametophyte in rice [J]. *J Exp Bot*, 2015, **66**(1): 99 – 112.
- [20] LI Haohao, LIU Xin, AN Jianping, *et al.* Cloning and elucidation of the functional role of apple *MdLBD13* in anthocyanin biosynthesis and nitrate assimilation [J]. *Plant Cell Tissue Organ Cult*, 2017, **130**(1): 47 – 59.
- [21] ZHANG Xueying, HE Yuqing, HE Wenda, *et al.* Structural and functional insights into the LBD family involved in abiotic stress and flavonoid synthases in *Camellia sinensis* [J]. *Sci Rep*, 2019, **9**: 14. doi: [10.1038/s41598-019-52027-6](https://doi.org/10.1038/s41598-019-52027-6).
- [22] LIU Hengzhi, CAO Minxuan, CHEN Xiaoli, *et al.* Genome-wide analysis of the lateral organ boundaries domain (LBD) gene family in *Solanum tuberosum* [J]. *Int J Mol Sci*, 2019, **20**(21): 5360. doi: [10.3390/ijms20215360](https://doi.org/10.3390/ijms20215360).
- [23] LU Qiang, SHAO Fenjuan, MACMILLAN C, *et al.* Genomewide analysis of the lateral organ boundaries domain gene family in *Eucalyptus grandis* reveals members that differentially impact secondary growth [J]. *Plant Biotechnol J*, 2018, **16**(1): 124 – 136.
- [24] HUANG Xiaoyun, LIU Gang, ZHANG Weiwei. Genome-wide analysis of LBD (lateral organ boundaries domain) gene family in *Brassica rapa* [J]. *Braz Arch Biol Technol*, 2018, **61**: e18180049. doi: [10.1590/1678-4324-2018180049](https://doi.org/10.1590/1678-4324-2018180049).
- [25] GRIMPLET J, PIMENTEL D, AGUDELO-ROMERO P, *et al.* The *LATERAL ORGAN BOUNDARIES* Domain gene family in grapevine: genome-wide characterization and expression analyses during developmental processes and stress responses [J]. *Sci Rep*, 2017, **7**: 15968. doi: [10.1038/s41598-017-16240-5](https://doi.org/10.1038/s41598-017-16240-5).
- [26] YANG Hui, SHI Guixia, DU Hongyang, *et al.* Genome-wide analysis of soybean *LATERAL ORGAN BOUNDARIES* domain-containing genes: a functional investigation of *GmLBD12* [J]. *Plant Genome*, 2017, **10**(1). doi: [10.3835/plantgenome2016.07.0058](https://doi.org/10.3835/plantgenome2016.07.0058).
- [27] HUANG Youjun, XIAO Lihong, ZHANG Zhongren, *et al.* The genomes of pecan and Chinese hickory provide insights into *Carya* evolution and nut nutrition [J]. *GigaScience*, 2019, **8**(5). doi: [10.1093/gigascience/giz036](https://doi.org/10.1093/gigascience/giz036).
- [28] GUO Wenlei, CHEN Junhao, LI Jian, *et al.* Portal of Juglandaceae: a comprehensive platform for Juglandaceae study [J].

- Hortic Res*, 2020, 7: 35. doi: [10.1038/s41438-020-0256-x](https://doi.org/10.1038/s41438-020-0256-x).
- [29] EL-GEBALI S, MISTRY J, BATEMAN A, *et al*. The Pfam protein families database in 2019 [J]. *Nucleic Acids Res*, 2019, 47(D1): D427 – D432.
- [30] FINN R D, CLEMENTS J, EDDY S R. HMMER web server: interactive sequence similarity searching [J]. *Nucleic Acids Res*, 2011, 39(S2): W29 – W37.
- [31] WANG Yupeng, TANG Haibao, DEBARRY J D, *et al*. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity [J]. *Nucleic Acids Res*, 2012, 40(7): e49. doi: [10.1093/nar/gkr1293](https://doi.org/10.1093/nar/gkr1293).
- [32] LARKIN M A, BLACKSHIELDS G, BROWN N P, *et al*. Clustal W and Clustal X version 2.0 [J]. *Bioinformatics*, 2007, 23(21): 2947 – 2948.
- [33] EDGAR R C. MUSCLE: multiple sequence alignment with high accuracy and high throughput [J]. *Nucleic Acids Res*, 2004, 32(5): 1792 – 1797.
- [34] PRICE M N, DEHAL P S, ARKIN A P. FastTree 2-approximately maximum-likelihood trees for large alignments [J]. *PLoS One*, 2010, 5(3): e9490. doi: [10.1371/journal.pone.0009490](https://doi.org/10.1371/journal.pone.0009490).
- [35] JONES D T, TAYLOR W R, THORNTON J M. The rapid generation of mutation data matrices from protein sequences [J]. *Comput Appl Biosci*, 1992, 8(3): 275 – 282.
- [36] SUBRAMANIAN B, GAO Shenghan, LERCHER M J, *et al*. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees [J]. *Nucleic Acids Res*, 2019, 47(W1): W270 – W275.
- [37] HU Bo, JIN Jinpu, GUO Anyuan, *et al*. GSDB 2.0: an upgraded gene feature visualization server [J]. *Bioinformatics*, 2015, 31(8): 1296 – 1297.
- [38] SCHMUTZ J, CANNON S B, SCHLUETER J, *et al*. Genome sequence of the palaeopolyploid soybean [J]. *Nature*, 2010, 463(7278): 178 – 183.
- [39] GUAN Rui, ZHAO Yunpeng, ZHANG He, *et al*. Draft genome of the living fossil *Ginkgo biloba* [J]. *GigaScience*, 2016, 5(1): 49. doi: [10.1186/s13742-016-0154-1](https://doi.org/10.1186/s13742-016-0154-1).
- [40] CONANT G C, WOLFE K H. Turning a hobby into a job: how duplicated genes find new functions [J]. *Nat Rev Genet*, 2008, 9(12): 938 – 950.
- [41] PRINCE V E, PICKETT F B. Splitting pairs: the diverging fates of duplicated genes [J]. *Nat Rev Genet*, 2002, 3(11): 827 – 837.
- [42] CANNON S B, MITRA A, BAUMGARTEN A, *et al*. The roles of segmental and tandem gene duplication in the evolution of large gene families in *Arabidopsis thaliana* [J]. *BMC Plant Biol*, 2004, 4: 10. doi: [10.1186/1471-2229-4-10](https://doi.org/10.1186/1471-2229-4-10).
- [43] LEE H W, KIM M J, PARK M Y, *et al*. The conserved proline residue in the LOB domain of LBD18 is critical for DNA-binding and biological function [J]. *Mol Plant*, 2013, 6(5): 1722 – 1725.
- [44] ARIEL F D, DIET A, CRESPI M, *et al*. The LOB-like transcription factor MtLBD1 controls *Medicago truncatula* root architecture under salt stress [J]. *Plant Signal Behav*, 2010, 5(12): 1666 – 1668.
- [45] 卢翔, 李效文, 夏海涛, 等. 薄壳山核桃胚发育过程中主要营养成分的变化 [J]. *林业科技*, 2018, 43(3): 40 – 44.
LU Xiang, LI Xiaowen, XIA Haitao, *et al*. Embryonic developmental and main nutritional components of *Carya illinoensis* [J]. *For Sci Technol*, 2018, 43(3): 40 – 44.
- [46] 刘力, 龚宁, 夏国华, 等. 山核桃种仁蛋白质及氨基酸成分含量的变异分析 [J]. *林业科学研究*, 2006, 19(3): 376 – 378.
LIU Li, GONG Ning, XIA Guohua, *et al*. Analysis on the protein and amino acids contents in hickory (*Carya cathayensis* sarg.) kernel and their variations [J]. *For Res*, 2006, 19(3): 376 – 378.
- [47] 于敏, 徐宏化, 王正加, 等. 6 个薄壳山核桃品种的形态及营养成分分析 [J]. *中国粮油学报*, 2013, 28(12): 74 – 77.
YU Min, XU Honghua, WANG Zhengjia, *et al*. Analysis of morphology and main nutrient components of 6 pecan varieties [J]. *J Chin Cereals Oils Assoc*, 2013, 28(12): 74 – 77.