

DOI: 10.11931/guihaia.gxzw201601026

赵艳, 翟莹, 宫国强, 等. 大豆 *PM34* 基因生物信息学预测及表达分析 [J]. 广西植物, 2016, 36(10):1220-1224

ZHAO Y, ZHAI Y, GONG GQ, et al. Bioinformatics and the expression pattern analysis of soybean *PM34* gene [J]. *Guihaia*, 2016, 36(10):1220-1224

大豆 *PM34* 基因生物信息学预测及表达分析

赵艳^{1*}, 翟莹¹, 宫国强², 王曼¹, 赵阳¹

(1. 齐齐哈尔大学 生命科学与农林学院, 黑龙江 齐齐哈尔 161006; 2. 吉林省食品检验所, 长春 130000)

摘要: 以大豆叶片总 RNA 为模板, 采用 RT-PCR 法从吉豆 2 号品种中克隆获得大豆种子成熟蛋白 (*PM34*) 基因序列, 利用生物信息学方法对大豆 *PM34* 基因编码的蛋白进行预测。结果表明: 该基因编码蛋白理论分子量为 31.7 kDa, 等电点为 6.60, 为亲水性蛋白; 该蛋白中无跨膜结构; 该蛋白中不存在信号肽。 *PM34* 基因编码蛋白的二级结构中 α 螺旋占 12.97%, 无规则卷曲占 41.30%, β 折叠占 45.73%。 *PM34* 基因编码蛋白的三级结构预测表明, 同源模建的模板是 3ijr.1.A, 是一种短链脱氢酶/还原酶, 与该蛋白的同源性为 54.65%。在进化关系上, 与绿豆、苜蓿的亲缘关系相对较近。采用实时定量 PCR 方法 (qRT-PCR), 检测大豆 *PM34* 基因在大豆各器官中的表达方式, 结果表明该基因在大豆根、茎、叶、花中的表达活性低, 而在种子中, 尤其是成熟种子中的相对表达活性很高。该研究结果为大豆 *PM34* 基因结构和功能的进一步研究奠定了基础。

关键词: 大豆, *PM34* 基因, 克隆, 生物信息学, qRT-PCR

中图分类号: Q943.2 文献标识码: A 文章编号: 1000-3142(2016)10-1220-05

Bioinformatics and the expression pattern analysis of soybean *PM34* gene

ZHAO Yan^{1*}, ZHAI Ying¹, GONG Guo-Qiang²,
WANG Man¹, ZHAO Yang¹

(1. College of Life Sciences and Agroforestry, Qiqihaer University, Qiqihaer 161006, Heilongjiang, China;

2. Jilin Provincial Institute of Food Inspection, Changchun 130000, China)

Abstract: In this paper, the structure and tissue-specificity of the soybean seeds mature protein gene (*PM34*) were studied, which could become the the foundations for further research the function of *PM34* gene in soybean or the other plants. The sequence of soybean(Jidou2) *PM34* gene was amplified by RT-PCR with the total RNA of soybean leaf as templet. The structure of soybean *PM34* gene was predicted by bioinformatics. The expression pattern of soybean *PM34* gene in soybean organs was detected by real-time quantitative PCR (qRT-PCR). The results showed that the molecular weight of the protein was 31.7 kDa, the isoelectric point was 6.60, and belonging to the hydrophilic protein. There was absent transmembrane domain and singal peptide. The secondary structure was composed of 12.97% alpha helix, 41.30% coil and 45.73% extended strand. The protein tertiary structure prediction showed that the homologous model template was 3 ijr. 1. A, which was a kind of short chain dehydrogenase/reductase and the homology was 54%. The evolutionary relationship demonstrated that soybean had relatively closer relationship with *Vigna radiata* and *Medicago trun-*

收稿日期: 2016-03-17 修回日期: 2016-09-15

基金项目: 黑龙江省自然科学基金(C201458); 黑龙江省普通本科高等学校青年创新人才培养计划(UNPYSCT_2016090); 国家自然科学基金(31301335) [Supported by the Natural Science Foundation of Heilongjiang Province (C201458); Training Plan of Young Creative Talents of Average Four-Year College of Heilongjiang (UNPYSCT_2016090); the National Nature Science Foundation of China (31301335)].

作者简介: 赵艳(1981-), 女, 吉林吉林市人, 博士, 副教授, 研究方向为植物分子生物学, (E-mail) zhaoyan3053877@163.com。

* 通讯作者

catula than others. In addition, the result showed that there was a little activity in roots, stems, leaves and flowers, but there was the highest activity in seeds, especially in mature seeds.

Key words: soybean, *PM34* gene, cloning, bioinformatics, qRT-PCR

大豆种子中含有丰富的蛋白质, 蛋白氨基酸组分齐全, 其具有抗肿瘤, 预防心血管疾病, 增强免疫力等保健作用 (Min et al, 2015)。大豆种子发育过程主要有 3 个时期, 即形态建成期, 成熟期和脱水干燥期。在种子成熟期阶段, 淀粉、蛋白质、脂肪等物质在种子中大量积累 (Kim & Lim, 2014), 其中蛋白质占 40% 左右, 种子贮藏蛋白在种子形成过程中逐渐合成和积累, 是大豆种子萌发中氨基酸和氮的主要来源 (刘春等, 2008)。种子蛋白包括贮藏蛋白、结构蛋白、酶蛋白及其它功能性蛋白等, 其中贮藏蛋白占种子蛋白的 70%~80%。根据离心时具有的不同沉降值将大豆贮藏蛋白分为 2S、7S、11S 和 15S 4 种组分 (Singh et al, 2015; Rudakova et al, 2014), 其它蛋白还包括脉酶、植物凝集素和胰蛋白酶抑制剂等 (Meyerl et al, 2012)。颜克亮等 (2008) 分析大豆种子形态建成期与成熟期基因差异表达情况, 获得在发育过程中可能影响大豆种子形态建成及蛋白质、油脂合成代谢等调控的重要基因, 检测表明大豆种子成熟蛋白基因 (*PM34*) 在大豆种子中表达量较高 (颜克亮等, 2008), 但并未对 *PM34* 基因编码蛋白的结构, 该基因在大豆各器官中的表达量做进一步研究, 此外, 至今在各种植物中对 *PM34* 基因的研究都鲜有报道。

本研究主要是克隆大豆 *PM34* 基因的序列, 利用生物信息学分析该基因编码蛋白的一级、二级、三级结构及系统进化关系; 此外, 通过 qRT-PCR 方法检测大豆 *PM34* 基因在大豆各器官中的表达情况, 为大豆 *PM34* 基因结构功能研究及其他植物中 *PM34* 基因的研究奠定基础。

1 材料与方 法

1.1 材料及主要试剂

大豆品种吉豆 2 号为本实验室保存; 大肠杆菌 DH 5 α 感受态购自北京索莱宝公司; TRIzol reagent 购自 Invitrogen 公司; Oligo (dT)₁₈、M-MLV 反转录酶、Recombinant RNase Inhibitor、ExTaq、pMD18-T 克隆载体、SYBP Premix Ex Taq II 荧光定量试剂盒、限制性内切酶 *Hind* III 和 *Kpn* I 均购自 Takara 公司;

DNA 凝胶回收试剂盒购自维特洁公司; 引物合成、测序由上海生工生物工程公司完成。

1.2 大豆 *PM34* 基因序列的克隆

取大豆 4 叶期的叶片, 提取总 RNA, 按照逆转录步骤合成大豆 cDNA 的第一条链。根据大豆基因组序列 (<http://www.phytozome.net/soybean>), 利用 Primer Premier 5.0 软件, 设计引物, (Z1: 5'-ATG-GCTTCCGGTGAACAGA-3'; Z2: 5'-TTAACCATTCACAACGGTCCAC-3'), RT-PCR 扩增大豆 *PM34* 基因序列。反应条件为预变性 94 $^{\circ}$ C 5 min, 94 $^{\circ}$ C 30 s, 58 $^{\circ}$ C 40 s, 72 $^{\circ}$ C 1 min, 共 30 个循环; 72 $^{\circ}$ C 后延伸 10 min。0.8% 琼脂糖凝胶电泳检测扩增产物, 凝胶回收试剂盒回收扩增片段, 按照试剂盒说明将扩增片段与 pMD18-T 克隆载体连接, 转化大肠杆菌 DH 5 α 感受态, 摇菌, 重组质粒双酶切鉴定后测序。

1.3 大豆 *PM34* 基因编码蛋白的生物信息学分析

利用 TMHMM Serverv 2.0 和 ExPASy Proteomics Server ([http://expasy.org.tools/protparam.html](http://expasy.org/tools/protparam.html)) 在线软件, 预测大豆 *PM34* 基因编码蛋白的一级结构: 基本理化性质和跨膜结构 (王同淑等, 2015); 利用 SignalP4.0 和 PORTER 在线软件, 预测大豆 *PM34* 基因编码蛋白的二级结构: 信号肽和蛋白质二级结构 (许瑞瑞等, 2012); 利用 SWISS-MODLE 在线软件, 预测大豆 *PM34* 基因编码蛋白的三级结构 (郑玲等, 2011)。

通过 NCBI 中 BLAST, 对大豆 *PM34* 基因进行氨基酸的同源性搜索。利用多序列比对软件 ClustalX 和 MEGA 5.0 软件, 采用邻接法 (Neighbor Joining), 1 000 次自检举 (bootstrapping), 构建系统发育树 (Sun et al, 2014; Saitou & Nei, 1987; Kumar et al, 2001; Yuan et al, 2014)。

1.4 QTR-PCR

分别提取大豆根、茎、叶、花、未成熟种子 (开花后 30 d)、成熟种子 (开花后 90 d) 的总 RNA, 按照逆转录的步骤分别合成 cDNA 的第一条链。以上不同大豆器官、不同发育阶段的材料 cDNA 为模板, 进行 qRT-PCR 反应。大豆的内参基因为 β -tubulin (GMU12286), 上、下游引物分别为 (B₁: 5'-GGAAG-GCTTTCTTGCATTGGTA-3'; B₂: 5'-AGTGGCATCCT-

GGTACTGC-3') (Zhao et al, 2015)。根据大豆 *PM34* 基因的核苷酸序列,利用 Primer Premier 5.0 软件设计检测大豆 *PM34* 基因在大豆各器官中表达量的引物,上游引物为 Y_1 :5'-TGCGAAACTATTG-GACTACACG-3',下游引物为 Y_2 :5'-GAAAGAGGCT-GGTATCAAAGG-3'。反应循环参数为 95 °C 10 s,58 °C 20 s,72 °C 30 s (Pan et al, 2015; Acharya et al, 2014; Schmidt & Delaney, 2010)。

2 结果与分析

2.1 大豆 *PM34* 基因序列的克隆

根据大豆基因组中的序列,RT-PCR 方法克隆获得大豆 *PM34* 基因核苷酸序列,长为 882 bp 的条带(图 1)。将该片段连接到 pMD18-T 载体上,获得重组质粒 pMD18-T-*PM34*。从大肠杆菌菌液中提取质粒,用 *Hind*III 和 *Kpn*I 进行双酶切鉴定,得到与预期相符的片段(图 1)。利用 DNAMAN 软件,将测序结果与大豆基因组中该段序列进行同源性比对,同源性达到 100%。

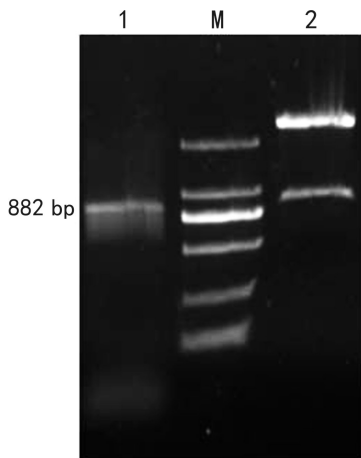


图 1 PCR 扩增产物及 pMD18-T-*PM34* 酶切鉴定
M. 2 000 分子标识; 1. PCR 扩增产物; 2. pMD18-T-*PM34* 酶切鉴定。
Fig. 1 PCR amplification and restriction enzyme digestion identification of pMD18-T-*PM34* M. 2 000 marker; 1. The PCR product; 2. *Hind* III and *Kpn* I digestion of pMD18-T-*PM34*.

2.2 大豆 *PM34* 基因的生物信息学预测

2.2.1 大豆 *PM34* 基因编码蛋白的一级结构预测

利用 ProtParam 在线软件进行蛋白质基本理化性质分析,预测得知大豆 *PM34* 基因编码蛋白的分子式为 $C_{1404}H_{2213}N_{385}O_{431}S_{12}$, 分子量为 31.7 KD, 等电点

(pI) 为 6.60, 带正电残基 (Arg + Lys) 总数为 30, 带负电残基 (Asp + Glu) 总数为 31, 总的亲水性平均系数为 -0.286, 该蛋白属于亲水性蛋白。氨基酸组成见表 1。此外,通过 TMHMM 在线软件对大豆 *PM34* 基因编码蛋白进行跨膜区预测,结果表明该蛋白无跨膜结构。

2.2.2 大豆 *PM34* 基因编码蛋白的二级结构预测

利用 SignalP4.0 软件预测得知,大豆 *PM34* 基因编码的蛋白不存在信号肽;用 PORTER 在线软件预测蛋白二级结构表明,此蛋白的 α 螺旋占 12.97%, 无规则卷曲占 41.30%, β 折叠占 45.73%。

2.2.3 大豆 *PM34* 基因编码蛋白的三级结构预测

利用 SWISS-MODEL 在线软件预测表明,同源模建的模板是 3ijr.1.A, 它是一种短链脱氢酶/还原酶,与大豆 *PM34* 基因编码蛋白的同源性为 54.65%。预测的三级结构如图 2。

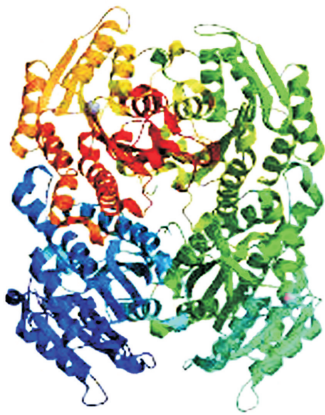
2.2.4 系统进化树分析 除有报道该基因为大豆种子成熟蛋白基因外,在各种植物中对 *PM34* 基因的研究都很少。本研究在 NCBI 中搜索与大豆 *PM34* 基因编码氨基酸序列同源性高的序列,分别来源于大豆 (*Glycine max*)、绿豆 (*Vigna radiata*)、菜豆 (*Phaseolus vulgaris*)、苜蓿 (*Medicago truncatula*)、桑树 (*Morus notabilis*) 和拟南芥 (*Arabidopsis*)。利用 ClustalX 和 Mega5.0 软件,构建大豆 *PM34* 基因编码氨基酸序列的系统进化树,进化关系如图 3 所示:大豆 *PM34* 基因 (AAF89645) 与大豆葡萄糖、核糖醇脱氢酶基因 (NP 001242815) 关系最近,同源性为 95.56%。其次是绿豆葡萄糖、核糖醇脱氢酶基因 (XP 014513287),苜蓿葡萄糖、核糖醇脱氢酶基因 (XP 003591094),同源性分别为 89.42%、88.40%。而与拟南芥葡萄糖、核糖醇脱氢酶基因 (NP 564670.2) 及拟南芥氧化还原酶基因 (XP 002891915) 的亲缘关系最远,同源性分别为 67.16%、64.37%。以上结果比较符合传统的分类系统,大豆、绿豆、苜蓿同属双子叶植物中的豆科植物,拟南芥属于单子叶植物中的十字花科植物。

2.3 大豆 *PM34* 基因的表达特性

大豆持家基因 β -tubulin 为内参,利用 qRT-PCR 方法对大豆根、茎、叶、花、未成熟种子和成熟种子中 *PM34* 基因的相对表达量进行分析。图 4 结果显示,大豆 *PM34* 基因在大豆根、茎、叶、花中的表达低,相对于叶中的表达量,大豆 *PM34* 基因在根、茎、花中表达活性分别为 10.4、0.08、0.49;而在大豆种子

表 1 大豆 *PM34* 基因编码蛋白的氨基酸组成Table 1 Amino acid composition of encoded protein by *PM34* gene in soybean

氨基酸 Amino acid	比例 Ratio (%)	氨基酸 Amino acid	比例 Ratio (%)	氨基酸 Amino acid	比例 Ratio (%)	氨基酸 Amino acid	比例 Ratio (%)
丙氨酸 Ala (A)	9.9	谷氨酰胺 Gln (Q)	5.1	亮氨酸 Leu (L)	5.8	丝氨酸 Ser (S)	6.1
精氨酸 Arg (R)	3.8	谷氨酸 Glu (E)	5.8	赖氨酸 Lys (K)	6.5	苏氨酸 Thr (T)	6.5
天冬酰胺 Asn (N)	4.1	甘氨酸 Gly (G)	8.5	甲硫氨酸 Met (M)	2.4	色氨酸 Trp (W)	0.3
天冬氨酸 Asp (D)	4.8	组氨酸 His (H)	2.0	苯丙氨酸 Phe (F)	3.8	酪氨酸 Tyr (Y)	3.8
半胱氨酸 Cys (C)	1.7	异亮氨酸 Ile (I)	5.5	脯氨酸 Pro (P)	5.8	缬氨酸 Val (V)	7.8

图 2 大豆 *PM34* 基因编码蛋白的三级结构预测Fig. 2 Protein tertiary structure prediction of soybean *PM34* proteins

中,尤其是成熟种子中的相对表达活性很高,相对于叶中的表达量,大豆 *PM34* 基因在未成熟种子和成熟种子中的表达活性分别为 219.79 和 930.65。

3 讨论与结论

PM34 基因是大豆种子成熟蛋白基因,其编码蛋白可能在种子发育过程中起作用。*PM34* 基因在各种植物中鲜有报道。本研究提取大豆吉豆 2 号品种的总 RNA,克隆大豆 *PM34* 基因的序列,测序结果表明克隆的 *PM34* 基因序列与大豆基因组 (williams 82) 中的相应序列一致,说明品种间无差异。通过生物信息学相关软件对该基因编码蛋白的理化性质、信号肽、跨膜结构域、蛋白质二级结构、蛋白质三级结构及系统进化等方面进行预测,这些结果有助于该基因所编码蛋白的功能研究。

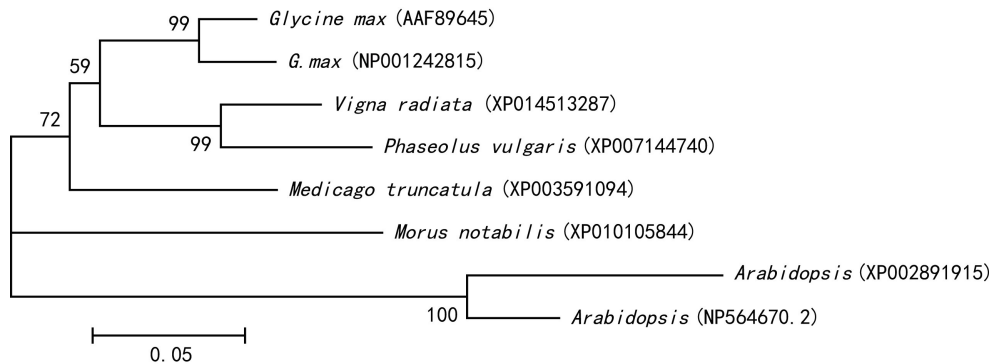
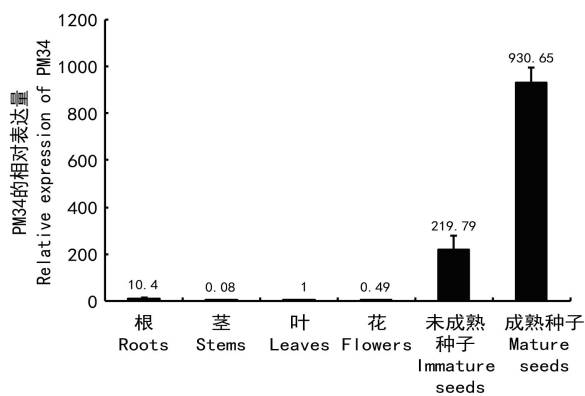
本研究中根据大豆 *PM34* 基因编码的氨基酸序列构建系统进化树,在 NCBI 的 BLAST 中进行大豆

PM34 基因编码蛋白序列的搜索并构建进化树。无论是同源性达到 95.56% 的来源于大豆的基因,还是绿豆、苜蓿中高同源性的基因,甚至是进化关系最远的拟南芥中的基因,这些基因都是葡萄糖、核糖醇脱氢酶基因,说明大豆 *PM34* 基因很可能是另一个大豆葡萄糖、核糖醇脱氢酶基因,该方面仍需进一步的实验验证。

由 qRT-PCR 结果可知,大豆 *PM34* 基因的表达主要集中在种子中,在其他器官中的相对表达量较低,属于种子特异表达基因,这可能与它是种子成熟蛋白基因有关,此外,其上游的启动子很可能具有调控下游基因在种子中特异表达的特性 (Yang et al, 2012; Fu et al, 2010; Zavallo et al, 2010)。本研究为大豆 *PM34* 基因及其上游启动子的结构和功能研究提供理论基础。大豆 *PM34* 基因是否为另一个大豆葡萄糖、核糖醇脱氢酶基因,*PM34* 基因及其启动子的功能研究将是下一步实验内容。

参考文献:

- ACHARYA S, RANJAN R, PATTANAIK S, et al, 2014. Efficient chimeric plant promoters derived from plant infecting viral promoter sequences [J]. *Planta*, 239:381-396.
- FU G, ZHONG Y, LI C, et al, 2010. Epigenetic regulation of peanut allergen gene *Arah3* in developing embryos [J]. *Planta*, 231 (5):1049-1060.
- KIM E, LIM UT, 2014. Effect of food deprivation period on the development and reproduction of *riptortus pedestris* (Hemiptera: Alydidae), and its egg parasitism [J]. *J Econ Entomol*, 107 (5):1785-1791.
- KUMAR S, TAMARA K, JAKOBSEN IB, et al, 2001. MEGA2: molecular evolutionary genetics analysis software [J]. *Bioinformatics*, 17:1244-1245.
- LIU C, WANG XS, MA H, 2008. Genetic improvement on soybean seed storage proteins [J]. *Soyb Sci*, 27(5): 866-872. [刘春, 王显生, 麻浩, 2008. 大豆种子贮藏蛋白遗传改良研究进展 [J]. *大豆科学*, 27(5):866-872.]
- MEYER LJ, GAO J, XU D, et al, 2012. Phosphoproteomic analysis

图3 大豆 *PM34* 基因编码氨基酸的进化树分析Fig. 3 Phylogenetic relationships of soybean *PM34* gene图4 大豆 *PM34* 基因在大豆各器官中的表达Fig. 4 Expression of soybean *PM34* gene in different soybean organs

of seed maturation in arabidopsis, rapeseed, and soybean [J]. *Plant Physiol*, 159(1):517-528.

MIN CW, GUPTA R, KIM SW, et al, 2015. Comparative biochemical and proteomic analyses of soybean seed cultivars differing in protein and oil content [J]. *J Agr Food Chem*, 63(32): 7134-7142.

PAN YL, MA X, LIANG HW, et al, 2015. Spatial and temporal activity of the foxtail millet (*Setaria italica*) seed-specific promoter *pF128* [J]. *Planta*, 241:57-67.

RUDAKOVA AS, RUDAKOV SV, KAKHOVSKAYA IA, et al, 2014. 11S Storage globulin from pumpkin seeds; regularities of proteolysis by papain [J]. *Biochemistry (Mosc)*, 79(8): 820-825.

SAITOU N, NEI M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees [J]. *Mol Biol Evol*, 4:406-425.

SCHMIDT GW and DELANEY SK, 2010. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress [J]. *Mol Genet Genom*, 283:233-241.

SINGH A, MEENA M, KUMAR D, et al, 2015. Structural and

functional analysis of various globulin proteins from soy seed [J]. *Crit Rev Food Sci*, 55(11):1491-1502.

SUN L L, LI Y, LI SS, et al, 2014. Identification and characterization of DfCHS, a chalcone synthase gene regulated by temperature and ultraviolet in *Dryopteris fragrans* [J]. *Cell Mol Biol*, 60(6):1-7.

WANG TS, LI CF, CHEN ZY, et al, 2015. Cloning and bioinformatics analysis of Peking Duck retinoic acid inducible gene I [J]. *Acta Agric Zhejiang*, 27(5): 734-739. [王同淑,李传峰,陈宗艳,等,2015. 北京鸭 *RIG-I* 基因的克隆及其生物信息学分析 [J]. *浙江农业学报*,27(5): 734-739.]

XU RR, ZHANG SZ, CAO H, et al, 2012. Bioinformatics analysis of WRKY transcription factor genes family in apple [J]. *Acta Horti Sin*, 39(10): 2049-2060. [许瑞瑞,张世中,曹慧,等,2012. 苹果 WRKY 转录因子家族基因生物信息学分析 [J]. *园艺学报*,39(10): 2049-2060.]

YANG L, HIROSE S, TAKAHASHI H, et al, 2012. Recombinant protein yield in rice seed is enhanced by specific suppression of endogenous seed proteins at the same deposit site [J]. *Plant Biotechnol J*, 10(9):1035-1045.

YAN K L, CHEN B, YAO JL, et al, 2008. Analysis of differentially expressed genes at pattern formation and maturation stages of soybean seed [J]. *J Wuhan Univ: Nat Sci Ed*, 54(2): 202-208. [颜克亮,陈波,姚家玲,等,2008. 大豆种子形态建成期与成熟期正反馈抑制消减文库构建及差异表达基因分析 [J]. *武汉大学学报·理学版*,54(2): 202-208.]

YUAN Y, YU SL, YU J, et al, 2014. Prediction the function of 4-Coumarate:CoA Logase(LJ4CL 1) in *Lonicera japonica* [J]. *Int J Mol Sci*, 15: 2386-2399.

ZAVALLO D, LOPEZBILBAO M, HOPP HE, et al, 2010. Isolation and functional characterization of two novel seed-specific promoters from sunflower (*Helianthus annuus* L.) [J]. *Plant Cell Rep*, 29(3):239-248.

ZHAO Y, SHA W, WANG QY, et al, 2015. Molecular cloning and activity analysis of a seed-specific FAD2-1B gene promoter from *Glycine max* [J]. *Cell Mol Biol*, 61(4):85-89.

ZHENG L, HUO XL, WANG FR, et al, 2011. Bioinformatic analysis on two *GhBI-1* genes of upland cotton [J]. *Acta Agric Shangdong*, 3: 4-7. [郑玲,霍雪寒,王芙蓉,等,2011. 陆地棉两个 *GhBI-1* 基因的生物信息学分析 [J]. *山东农业学报*, 3: 4-7.]