

大豆 *GmIDD* 及其同源基因的生物信息学分析及功能预测

杨 雪,王 阔,李英华,郑艳红,姚愈恒,李敏敏,赵 琳,李文滨

(东北农业大学 大豆研究所/大豆生物学教育部重点实验室,黑龙江 哈尔滨 150030)

摘 要:利用 NCBI 和 Phytozome 数据库进行 *GmIDD* 基因的同源基因检索,获取该基因在大豆不同染色体中 5 个拷贝的序列信息。通过对功能结构域的分析结果显示 *GmIDD* 基因 5 个拷贝均含有 2 个 C₂H₂ 型(72 ~ 92 aa、114 ~ 142 aa)、2 个 C₂HC 型(149 ~ 169 aa、176 ~ 195 aa) 锌指结构域。*GmIDD* 基因 cDNA 序列全长为 1 227 bp,编码 408 个氨基酸。系统进化树分析结果显示 *GmIDD* 基因 5 个拷贝中除了 Glyma. 14g10940 和 Glyma. 17G228000 外,其它 IDD 蛋白间进化距离较远,表明不同大豆 *IDD* 基因在功能上可能存在一定差异。长日照处理下 *GmIDD* 表达量的分析结果显示 *GmIDD* 基因受短日照诱导表达,因此推测 *GmIDD* 可能参与大豆开花的调控过程。

关键词:大豆; *GmIDD* 基因; 锌指结构域; 短日照

中图分类号:S565. 1 **文献标识码:**A **DOI:**10. 11861/j. issn. 1000-9841. 2017. 04. 0525

Bioinformatics Analysis and Functional Prediction of Soybean *GmIDD* and Its Homologous Genes

YANG Xue, WANG Kuo, LI Ying-hua, ZHENG Yan-hong, YAO Yu-heng, LI Min-min, ZHAO Lin, LI Win-bin
(Soybean Research Institute, Northeast Agricultural University/Key Laboratory of Soybean Biology of Chinese Education Ministry, Harbin 150030, China)

Abstract: In this study, we used the NCBI and Phytozome databases to retrieve the *GmIDD* gene homologous gene, and obtained 5 copies of the gene in different chromosomes of soybean. We analyzed the functional domain of the 5 copies of the *GmIDD* gene. The results showed that the five copies of *GmIDD* gene contained two C₂H₂ type(72 – 92 aa, 114 – 142 aa), two C₂HC type(149 – 169 aa, 176 – 195 aa) zinc finger domain. The full sequence of *GmIDD* transcription factor cDNA was 1 227 bp, encoding 408 amino acids. Phylogenetic tree analysis showed that the evolution of other IDD proteins was distant except Glyma. 14g10940 and Glyma. 17G228000 in the five copies of *GmIDD*, which indicated that there might be some differences in the function of different soybean *IDD* genes. The results of *GmIDD* expression by long and short daylight showed that *GmIDD* gene was induced by short daylight and we speculate that *GmIDD* may be involved in the regulation of soybean flowering.

Keywords: Soybean; *GmIDD*; Zinc finger domain; Short days

大豆是短日照(SD)植物,对光周期反应十分敏感^[1-2],短日照能促进开花,长日照则抑制花芽生长^[3]。在研究植物光周期途径初期大豆就被作为经典的模式植物^[4]。地球上大豆的种植横跨 85 个纬度,不同的大豆品种种植范围北至温哥华,南到新西兰,范围很广。但是由于纬度的限制,每一个大豆品种可以种植的区域很狭窄,适应性有限^[5]。因此开花时间和成熟期就成为了大豆重要的农艺性状,这一农艺性状控制着大豆品种的产量、品质以及广适性。在不同地区的不同环境条件下对大豆栽培品种进行合理的种植,使该地区的耕地资源得到充分的利用,对增加大豆产量具有重要意义^[6]。因此,为了减轻育种压力,发掘控制大豆开

花与成熟的光周期新基因,进一步阐明该类基因参与大豆光周期效应的分子机制,则显得尤为重要。

前人研究表明,拟南芥 *IDD* 基因在成花诱导过程中具有重要作用^[7]。*idd8* 突变体表现出晚花表型^[8]。并且在糖介导的成花诱导过程中与光周期和碳水化合物变化相关^[9]。*IDD* 基因参与糖的运输和代谢,*AtIDD8* 通过调节糖信号传导来控制拟南芥光周期依赖的开花途径^[10];*Ehd2* 是 *ZmID1* 在水稻中的直系同源物,在短日照条件下能够诱导 *FT* 基因上调表达进而促进开花^[11];在玉米中首次发现了 *IDD* 转录因子家族的第一个成员 *INDETERMINATE1 (ID1)*,其在玉米花期调控中起着重要作用^[12],*ID1* 是一个控制玉米开花转换的转录调节因

收稿日期:2017-03-07
基金项目:黑龙江省普通本科高等学校青年创新人才培养计划(UNPYSCT-2015003);国家重点研发计划“七大农作物育种”重点专项(2016YFD0101005);转基因生物新品种培育科技重大专项(2016ZX08004-005);黑龙江省自然科学基金(C2015009);东北农业大学“学术骨干”(16XG03);东北农业大学 2016 年大学生 SIPT 计划(201610224141);“十二五”农村领域国家科技计划课题(2013AA102602)。

第一作者简介:杨雪(1988-),女,硕士,主要从事大豆生物技术研究。E-mail: a18745021404@163.com。
通讯作者:李文滨(1958-),男,教授,博导,主要从事大豆生物技术研究。E-mail: wenbinli@yahoo.com;
赵琳(1980-),女,研究员,博导,主要从事大豆生物技术研究。E-mail: zhaolinneau@126.com。

子。丧失功能的 ID1 突变体的纯合子 *id1* 能够延长生殖生长,延迟开花,偶尔会形成营养生长特征的花序。表达分析表明, *ID1* 的 mRNA 仅在未成熟的叶组织中表达,在茎尖分生组织中不表达。这表明, *ID1* 可能调节一个长距离叶衍生的花遗传信号^[13]。 *ID1* 基因编码的蛋白是植物特有的创始成员锌指蛋白家族。这个家族的成员包括 4 个锌指结构域,它们共同构成了 ID 结构域(IDD)。 *ID1* 蛋白和其它 IDD 蛋白的生化分析显示,4 个锌指基序中的两个锌指基序具有独特的 DNA 结合特性。

本研究利用大豆基因组数据库进行 *IDD* 基因信息挖掘,通过对 *GmIDD* 基因及其编码氨基酸的理化性质与结构域进行分析,并检测其在长短日照条件下 *GmIDD* 基因的表达量,为 IDD 转录因子参与大豆光周期信号传导途径的研究提供重要的理论基础。

1 材料与方法

1.1 材料

- 1.1.1 植物材料 大豆品种东农 42。
- 1.1.2 主要试剂 总 RNA 提取试剂盒和 cDNA 第一链合成试剂盒购自天根生物公司;SYBR Green 荧光定量试剂盒购自 Invitrogen 公司。

1.2 方法

- 1.2.1 序列来源 本课题组前期通过抑制性差减杂交 cDNA 文库筛选得到 *GmIDD* 基因序列^[14],在 NCBI 和 Phytozome 数据库中检索获得与 *GmIDD* 蛋白相似度较高的 *IDD* 基因序列,分别为 *GmIDD* 基因(Glyma. 14g10940)、在不同染色体上的其余 4 个拷贝(Glyma. 17G228000、 Glyma. 07G158200、 Glyma. 08G192300 和 Glyma. 15G024500),以及拟南芥、玉米、水稻、蓖麻、葡萄、红小豆、木豆、绿豆、羽扇豆、胡桃、莲、胡杨、烟草、菜豆和可可等物种的 *IDD* 基因序列。本试验针对 Glyma. 14g10940 基因序列进行研究。
- 1.2.2 序列分析 采用 ExPASy Proteomics Server (<http://auexpasy.org/tools/protparam.html>)在线预测 *GmIDD* 的氨基酸组成、相对分子量、理论等电点、蛋白质的亲疏水性及稳定性等理化性质;利用 SignalP4.1 程序 (<http://www.cbs.dtu.dk/services/SignalP/>) 和 TMpred Server 软件 (http://www.ch.embnet.org/software/TMPRED_form.html) 分别分析了 *GmIDD* 蛋白信号肽和跨膜螺旋区;用 PSORT (<http://www.psort.org.uk/>) 软件进行亚细胞定位预测;用 NCBI 的 protein blast 功能进行同源蛋白检索;用 SOPMA (<https://npsa-prabi.ibcp.fr/>

cgi-bin/secpred_sopma.pl 和 SWISS - MODEL 在线软件(<http://swissmodel.expasy.org/SWISS-MODEL.html>)分别预测了 *GmIDD* 蛋白的二级结构和三级结构;利用 NCBI 数据库 (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) 进行 *GmIDD* 同源基因的检索,利用 ClustalX2 和 MEGA7.0 软件进行氨基酸序列的比对并构建系统进化树。

1.2.3 定量 PCR 检测 将东农 42 种子种植于培养钵中,每钵一粒。在 25℃,长日照(LD)下培养至第二片三出复叶完全展开。挑选长势一致的大豆幼苗分别进行长日照(LD)与短日照(SD)处理,分别在处理后的第 0,1,2,4 天同一时间点对第二片三出复叶进行取样,提取大豆叶片中的总 RNA,反转录后进行荧光定量 PCR,检测长短日照条件下 *GmIDD* 基因的表达量变化。根据 *GmIDD* 基因的编码区序列设计定量 PCR 引物,q*GmIDD* - F: CTC-CAACAGTGGTAGAAGAAGCAA;q*GmIDD* - R: AGG-GAAGGAAGCTCAAAGAAGAAAG。以大豆 *GmACTIN4* 基因为内参设计引物,q*GmACTIN4* - F: GTGT-CAGCCATACTGTCCCCATTT; q*GmACTIN4* - R: GTTTC AAGCTCTTGCTCGTAATCA。qRT - PCR 反应体系(20 μL): cDNA 模板 1 μL, 2 × SuperReal PreMix 10 μL,正向引物(10 μmol · L⁻¹) 0.5 μL,正向引物(10 μmol · L⁻¹) 0.5 μL。反应程序:1. 94℃ 10 min;2. 94℃ 30 s, 60℃ 30 s, 72℃ 30 s, 35 个循环。

1.3 数据分析

利用 Bio - Rad CFX Manager 软件进行数据处理,采用 2^{-ΔΔCt}法进行基因表达量分析,SPSS 17.0 软件进行生物统计学分析。

2 结果与分析

- 2.1 *GmIDD* 基因的理化性质和蛋白结构分析预测
- 2.1.1 理化性质分析 大豆 *GmIDD* 转录因子编码氨基酸序列长度为 408 aa, cDNA 序列全长为 1 227 bp,利用 ExPASy 软件对氨基酸全长序列的分子量和等电点分析,结果显示:等电点理论值为 9.25,估计分子量 45.129 kDa。通过 ProtParam 软件分析 *GmIDD* 氨基酸组成成分见表 1,其中带正电荷的精氨酸(Arg)和赖氨酸(Lys)总数为 45,占 11%;带负电荷的天冬氨酸(Asp)和谷氨酸(Glu)总数为 32,占 7.8%;疏水氨基酸 157 个,占 38.5%;亲水氨基酸 204 个,占 49.9%,脂肪族氨基酸指数为 71.27,总平均亲水性为 -0.498,说明该蛋白具有较强的亲水性。*GmIDD* 蛋白分子由 6 290 个原子组成,其分子式为 C₁₉₆₇H₃₁₂₂N₅₇₆O₆₀₇S₁₈。稳定参数为 54.61,大于 40,因此该蛋白是不稳定蛋白。

表 1 GmIDD 蛋白的氨基酸组成成分

Table 1 Protein amino acid composition of GmIDD		
氨基酸组成	数量	百分比
Amino acid	Number	Percentage/%
丙氨酸 Ala(A) *	28	6.9
精氨酸 Arg(R) #	20	4.9
天冬酰胺 Asn(N)	19	4.7
天冬氨酸 Asp(D) #	14	3.4
半光氨酸 Cys(C)	11	2.7
谷氨酰胺 Gln(Q) #	20	4.9
谷氨酸 Glu(E) #	18	4.4
甘氨酸 Gly(G)	17	4.2
组氨酸 His(H) #	20	4.9
异亮氨酸 Ile(I) *	16	3.9
亮氨酸 Leu(L) *	38	9.3
赖氨酸 Lys(K) #	25	6.1
甲硫氨酸 Met(M) *	7	1.7
苯丙氨酸 Phe(F) *	14	3.4
脯氨酸 Pro(P) *	24	5.9
丝氨酸 Ser(S) #	55	13.5
苏氨酸 Thr(T) #	32	7.8
色氨酸 Trp(W) *	4	1.0
酪氨酸 Tyr(Y) *	8	2.0
缬氨酸 Val(V) *	18	4.4

#: 亲水氨基酸; *: 疏水氨基酸。
#: Hydrophilic amino acid; *: Hydrophobic amino acid.

2.1.2 GmIDD 蛋白结构分析预测 GmIDD 蛋白经信号肽和跨膜区分析显示该蛋白氨基酸序列不含

有信号肽和跨膜区,该蛋白不属于分泌性跨膜蛋白质。通过 PSORT 预测 GmIDD 蛋白主要定位于细胞核中。GmIDD 蛋白氨基酸序列结构功能域分析发现该蛋白含有 4 个锌指结构域:2 个 C₂H₂ 型(72 ~ 92 aa、114 ~ 142 aa),2 个 C₂HC 型(149 ~ 169 aa、176 ~ 195 aa),预测该蛋白属于 IDD 转录因子家族。二级结构分析表明:GmIDD 蛋白的二级结构中包含 24.26% 的 α 螺旋,8.82% 的 β -转角,17.40% 的伸展链和 49.51% 的不规则卷曲结构(图 1)。利用 SWISS-MODEL 在线预测了 GmIDD 蛋白的三级结构,该结构图直观地呈现了二级结构预测的结果(图 2)。

2.2 GmIDD 基因与同源基因的氨基酸序列比对
从 NCBI 数据库中检索到 4 个 *GmIDD* (Glyma.14G10940) 同源基因,分别为 Glyma.17G228000、Glyma.07G158200、Glyma.08G192300、Glyma.15G024500。氨基酸序列同源性分析结果表明:*GmIDD* 及其同源基因的氨基酸序列均含有 C₂H₂ 和 C₂HC 型锌指结构域并且与 Glyma.07G158200、Glyma.08G192300、Glyma.15G024500 IDD 序列同源性分别为 33.87%、37.03% 和 40%;Glyma.14G10940 与 Glyma.17G228000 同源性最高,高达 69.16%(图 3)。

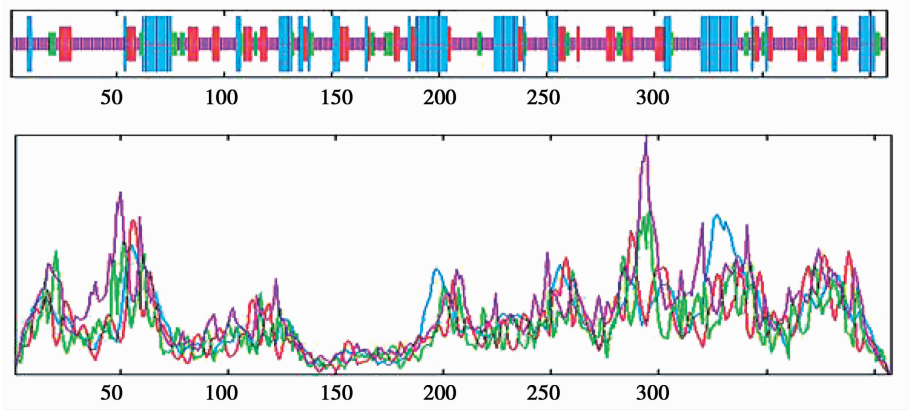


图 1 GmIDD 蛋白质二级结构
Fig. 1 Secondary structure of GmIDD protein

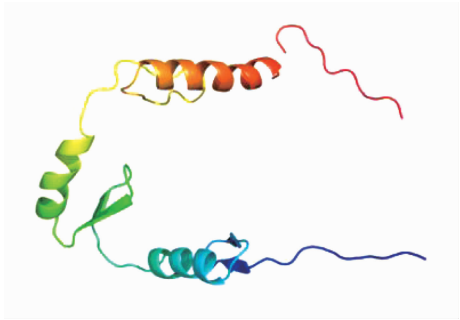
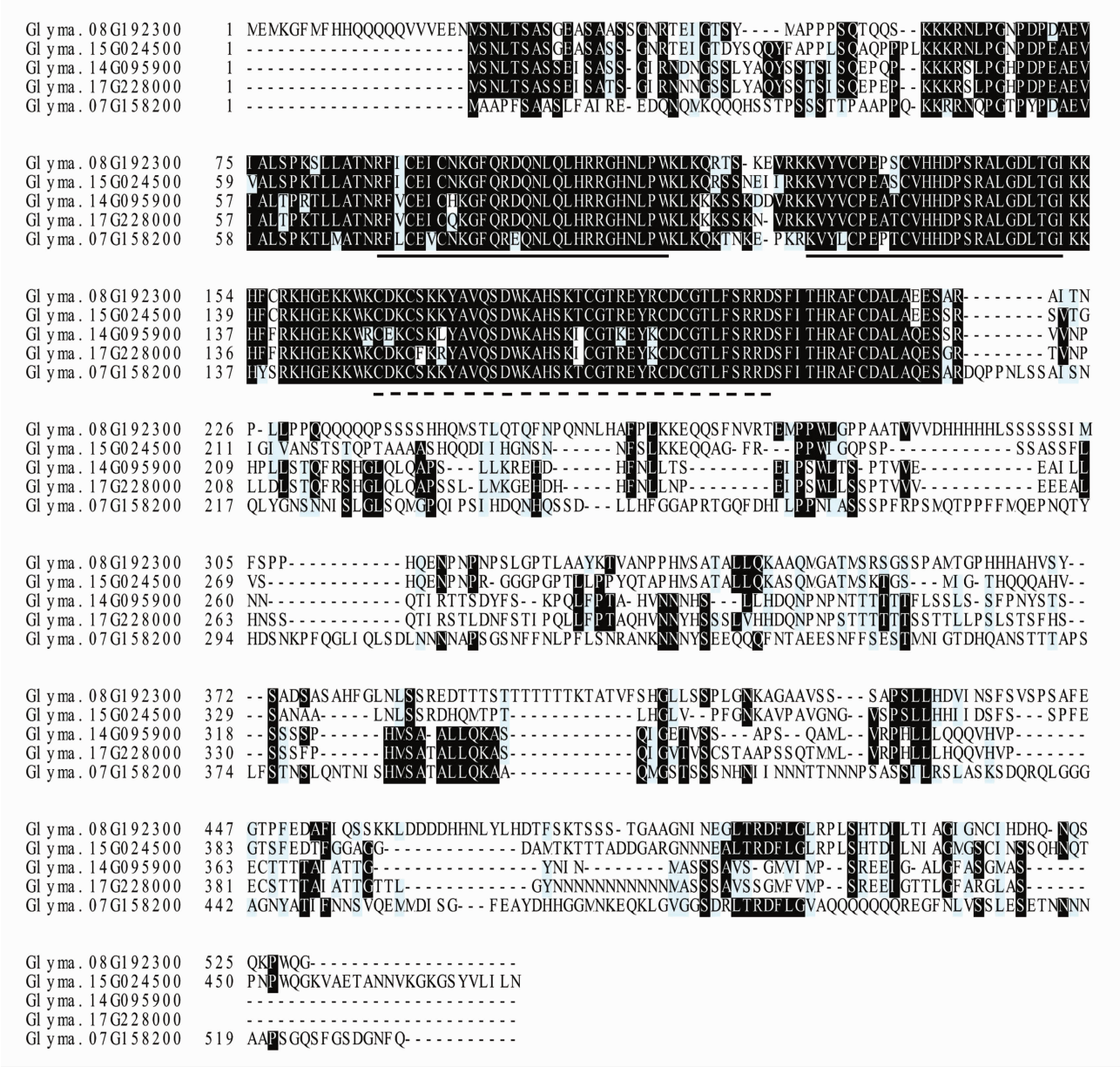


图 2 GmIDD 蛋白质三维结构
Fig. 2 3D structure of GmIDD protein

2.3 GmIDD 基因同源进化树分析
利用 Mega 7.0 软件构建了系统发生树(图 4),该发生树包含来自 16 个物种的 20 条 *IDD* 基因,结果表明 Glyma.14g10940 和 Glyma.17G228000 进化距离最近,该结果与同源基因的氨基酸序列比对结果一致,并且与菜豆(ESW33016)、红小豆(NC_030643.1)、木豆(KYP72220.1)、绿豆(XP_017428159.1)等豆科植物 *IDD* 蛋白位于同一分支上;Glyma.08G192300 与胡桃(XP_018848593.1)、烟草(NW_008911959.1)、莲(XP_010242234.1)以

及 Glyma. 15G024500 位于同一支上; Glyma. 07G158200 则与玉米 (GRMZM2G320287) 和水稻 (LOC_Os02g31890) 在同一支上。因此, *GmIDD* 基因的 5 个拷贝中除了 Glyma. 14g10940 与 Gly-

ma. 17G228000 外, 其它 *IDD* 蛋白进化距离较远, 表明大豆不同 *IDD* 同源基因在功能上可能存在一定差异。



C₂H₂ 型结构域由下划线标出; C₂HC 型结构域由下划线点标出。
C₂H₂ domains were underlined with lines; C₂HC domains were underlined dots respectively.

图 3 *GmIDD* 同源基因的氨基酸序列对比
Fig. 3 Amino acid sequence alignment of *GmIDD* homologous gene

2.4 长短日照下大豆 *GmIDD* 基因表达量变化

长势一致的大豆幼苗分别进行长日照 (LD) 与短日照 (SD) 处理。处理后的第 0、1、2、4 天对 *GmIDD* 基因的表达水平进行分析, 结果显示: 短日

照处理后的大豆 *GmIDD* 基因的表达量丰度显著高于在长日照 (LD) 条件下的表达丰度。说明 *GmIDD* 基因的表达受短日照诱导 (图 5)。

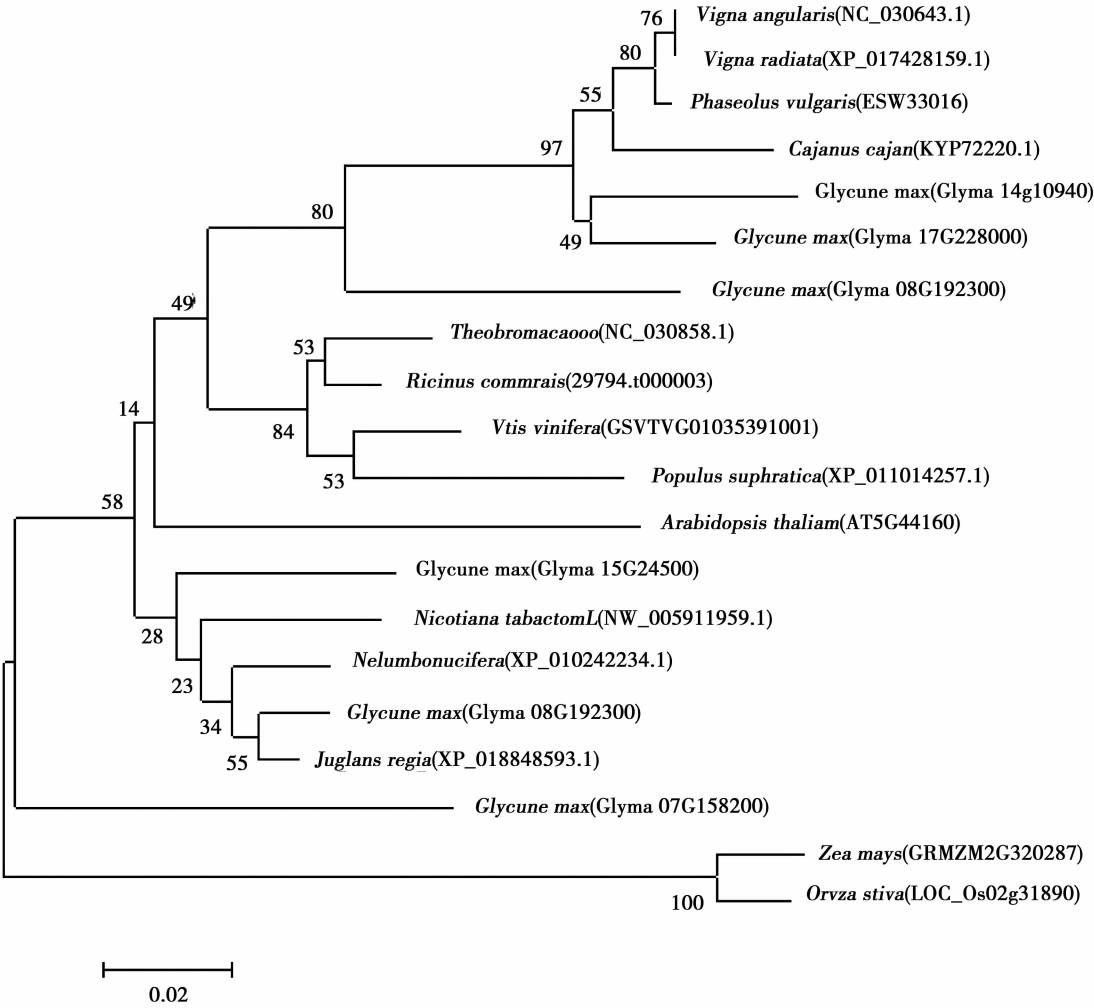


图 4 *GmIDD* 基因同源进化树分析

Fig. 4 Phylogenetic tree analysis of *GmIDD* gene

3 讨论

在玉米中首次发现了 *IDD* 转录因子家族的第一个成员 *INDETERMINATE1* (*ID1*), 其在玉米从营养生长转变到生殖生长中起到关键作用, *ID1* 基因编码的蛋白是植物特有的创始成员锌指蛋白家族, 这个家族的成员包括 4 个锌指结构域, 而本研究通过生物信息学分析发现 *GmIDD* 具有 4 个锌指结构域: 2 个 C_2H_2 型 (72 ~ 92 aa、114 ~ 142 aa), 2 个 C_2HC 型 (149 ~ 169 aa、176 ~ 195 aa), 进而预测该蛋白属于 *IDD* 转录因子家族; 氨基酸序列比对和系统进化树分析显示 *GmIDD* 基因的 5 个拷贝中除了 Glyma. 14g10940 与 Glyma. 17G228000 外, 其它 *IDD* 蛋白进化距离较远, 表明不同大豆 *IDD* 基因在功能上可能存在一定差异。 *GmIDD* 基因对长短日照下不同天数变化的响应分析显示 *GmIDD* 基因受短日照诱导表达, 推测 *GmIDD* 基因可能参与大豆光周期调控途径, 本研究应用生物信息学等相关技术手段对 *GmIDD* 蛋白的氨基酸结构、亲疏水性、进化树等

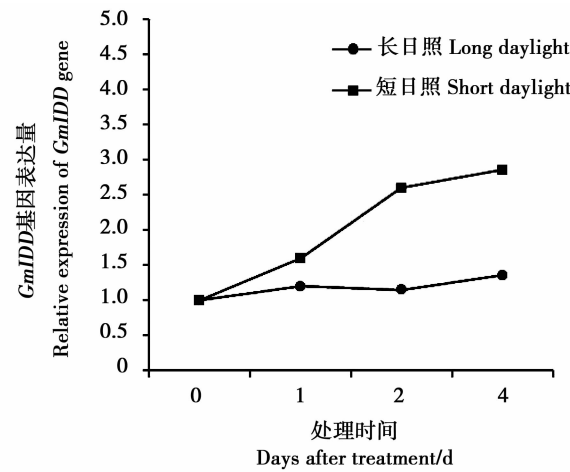


图 5 长短日条件下, 大豆 *GmIDD* 基因表达随处理时间的变化

Fig. 5 Change of expression of *GmIDD* with different treatment time under LDs and SDs

进行分析,并进一步阐明其在不同光照条件下的表达水平差异,为进一步解析 *GmIDD* 基因在大豆生长、发育和生殖过程中的作用提供重要的理论依据,但是 *GmIDD* 参与光周期调控的分子机理还有待于进一步研究。

4 结 论

理化性质分析结果显示大豆 *GmIDD* 转录因子编码氨基酸序列长度为 408 aa,cDNA 序列全长为 1 227 bp,等电点理论值为 9.25,估计分子量 45.129 kDa,总平均亲水性为 -0.498,*GmIDD* 蛋白分子式为 C₁₉₆₇H₃₁₂₂N₅₇₆O₆₀₇S₁₈,稳定参数为 54.61,属于不稳定蛋白;*GmIDD* 蛋白结构分析显示该蛋白的氨基酸序列含有 4 个锌指结构域:2 个 C₂H₂ 型(72~92 aa、114~142 aa),2 个 C₂HC 型(149~169 aa、176~195 aa),为大豆 IDD 转录因子家族成员,不含有信号肽和跨膜区,不属于分泌性跨膜蛋白质;*GmIDD* 基因同源进化树分析表明大豆不同 *IDD* 同源基因在功能上可能存在一定差异;短日照处理后的大豆 *GmIDD* 基因的表达量丰度显著高于在长日照(LD)条件下的表达丰度,说明 *GmIDD* 基因的表达受短日照诱导,推测 *GmIDD* 基因可能参与大豆光周期调控途径。

参考文献

[1] 任红玉. 中国大豆品种生育期性状及光温效应模型的研究 [D]. 哈尔滨: 东北农业大学, 2001. (Ren H Y. Study on growth dates characters and photo-thermal models of soybean varieties in China [D]. Harbin: Northeast Agricultural University, 2001.)

[2] 高振福, 单维奎. “铁丰 18 号”大豆易地种植的生态模式研究[J]. 中国油料作物学报, 1983, (3): 18-20. (Gao Z F, Shan W K. Study on ecological model of soybean planting in ‘Tiefeng 18’ [J]. Chinese Journal of Oil Crop Sciences, 1983; 18-20.)

[3] 吴辉. 大豆光周期相关基因 *GmMS* 的分离及功能初步分析 [D]. 海南: 海南大学, 2012. (Wu H. Cloning and preliminary functional analysis of the photoperiodic-related gene *GmMS* from soybean [D]. Hainan: Hainan University, 2012.)

[4] Zhang Q, Li H, Li R, et al. Association of the circadian rhythmic

expression of *GmCRY1a* with a latitudinal cline in photoperiodic flowering of soybean[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105 (52): 21028-21033.

[5] 张惠君, 王海英, 刘闯, 等. 不同来源亚有限型大豆品种适应性的比较研究[J]. 大豆科学, 2007, 26(3): 332-337. (Zhang H J, Wang H Y, Liu C, et al. Comparative studies on adaptability of semi-determine soybean varieties from different region[J]. Soybean Science, 2007, 26(3): 332-337.)

[6] 徐泽恒. 大豆光敏色素相互作用因子的筛选与功能鉴定 [D]. 北京: 中国科学院大学, 2014. (Xu Z H. The screening and identity of phytochrome interacting factors of soybean [D]. Beijing: University of Chinese Academy of Sciences, 2014.)

[7] Fan S, Zhang D, Xing L, et al. Phylogenetic analysis of *IDD* gene family and characterization of its expression in response to flower induction in *Malus*[J]. Molecular Genetics & Genomics, 2017; 1-17.

[8] Seo P J, Ryu J, Kang S K, et al. Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in *Arabidopsis*[J]. Plant Journal for Cell & Molecular Biology, 2011, 65(3): 418-429.

[9] Pagter M, Lefèvre I, Arora R, et al. Quantitative and qualitative changes in carbohydrates associated with spring deacclimation in contrasting *Hydrangea*, species[J]. Environmental & Experimental Botany, 2011, 72(3): 358-367.

[10] Funck D, Clauss K, Frommer W B, et al. The *Arabidopsis* CstF64-Like RSRI/ESPI protein participates in glucose signaling and flowering time control[J]. Frontiers in Plant Science, 2012, 3(4): 80.

[11] Matsubara K, Yamanouchi U, Wang Z X, et al. Ehd2, a rice ortholog of the maize *INDETERMINATE1* gene, promotes flowering by up-regulating *Ehd1* [J]. Plant Physiology, 2008, 148 (3): 1425.

[12] Wolfe S A, Nekludova L, Pabo C O. DNA recognition by Cys₂ His₂ zinc finger proteins [J]. Annual review of biophysics and biomolecular structure, 2000, 29: 183-212.

[13] Colasanti J, Yuan Z, Sundaresan V. The indeterminate, gene encodes a zinc finger protein and regulates a leaf-generated signal required for the transition to flowering in maize[J]. Cell, 1998, 93 (4): 593.

[14] 赵琳, 罗秋兰, 杨春亮, 等. 大豆在暗诱导条件下差异表达 cDNA 文库的构建及分析[J]. 大豆科学, 2007, 26(2): 134-139. (Zhao L, Luo Q L, Yang C L. Construction and analysis of differentially expressed cDNA library from soybeans induced by darkness [J]. Soybean Science, 2007, 26(2): 134-139.)