



大豆 *GmGW2* 基因的生物信息学分析

吕金东¹, 刘 灿², 屈亦潇², 陈汉谋², 杨翊晖², 丁 晨², 郭 蓓^{2,3}

(1. 北京农学院 植物科学技术学院, 北京 102206; 2. 北京农学院 生物科学与工程学院, 北京 102206; 3. 中华人民共和国农业农村部 华北都市农业重点实验室, 北京 102206)

摘 要: *GW2* 基因在多种植物中发挥调控籽粒性状的功能, 为预测分析大豆中 *GW2* 基因的功能, 前期通过同源克隆的方法从大豆中成功克隆出 *GmGW2* 基因的 cDNA 序列。为了继续开展深入研究, 本试验从大豆中黄 10 号中克隆出 *GmGW2* 基因的全长 DNA 序列并进行了较为详细的生物信息学分析。结果表明: *GmGW2* 全长 8 467 bp, 包含 8 个外显子和 7 个内含子, 编码 431 个氨基酸。预测是一个不稳定亲水蛋白, 定位于细胞核或细胞质中, 不含跨膜结构域及信号肽, 具有 Zinc finger 和 RING-type 结构域。二级、三级结构预测分析显示, α -螺旋及不规则卷曲是其整体蛋白质结构中的主要组成结构元件。利用该基因的 DNA 序列, 构建进化树分析得出该基因与木豆、相思子、鹰嘴豆亲缘关系较近。根据以上分析结果推测 *GmGW2* 基因可能也具有调控大豆籽粒发育的功能。此研究旨在为进一步挖掘和验证 *GmGW2* 基因功能奠定基础, 为大豆产量育种工作提供有益信息。

关键词: 大豆; 中黄 10 号; *GmGW2*; 生物信息学分析

Bioinformatics Analysis of Gene *GmGW2* in Soybean

LYU Jin-dong¹, LIU Can², QU Yi-xiao², CHEN Han-mou², YANG Yi-hui², DING Chen², GUO Bei^{2,3}

(1. College of Plant Science and Technology, Beijing University of Agriculture, Beijing 102206, China; 2. College of Biological Science and Engineering, Beijing University of Agriculture, Beijing 102206, China; 3. Key Laboratory of Urban Agriculture (North China), Ministry of Agriculture and Rural Affairs of the People's Republic of China, Beijing 102206, China)

Abstract: The *GW2* gene in diverse plants can control grain width and grain weight. In order to predicet and analysis the function of *GW2* gene in soybean, the cDNA sequence of the *GmGW2* gene was successfully cloned from soybean by the method of homologous cloning. In order to continue the in-depth study, the full-length DNA sequence of *GmGW2* gene was cloned from soybean Zhonghuang 10 and detailed bioinformatics analysis was carried out in this experiment. The results showed that the cloned *GmGW2* was consisted of 8 exons and 7 introns, encoding 431 amino acids, which was predicted to be located in the nucleus or cytoplasm, without transmembrane domain and signal peptide, and it was an unstable hydrophilic protein with a Zinc finger, RING-type domain. The secondary and third structure prediction analysis showed that α -helix and irregular curl were the main structural elements in the overall protein structure. Phylogenetic tree showed that the gene was closely related to *Cajanus cajan*, *Abrus precatorius* and *Cicer arietinum*. *GmGW2* gene may also play a regulational function in soybean seed traits above the results. The results of this study laid the foundation for further digging and verifying the function of *GmGW2* gene, aiming to provide useful information for soybean yield breeding.

Keywords: Soybean; Zhonghuang 10; *GmGW2*; Bioinformatics analysis

大豆是我国重要的粮食作物之一, 对环境适应性强, 含有丰富的蛋白质, 既可用作食用油原材料, 也可用作喂养动物的饲料, 因此, 大豆的育种与栽培在我国备受重视^[1]。在我国大豆种植面积无法大量增加的前提下, 为保证国内市场的大豆供应就必须提高大豆单位面积产量^[2]。因此, 提高大豆产量成为大豆育种工作的重中之重。

百粒重是大豆产量的重要构成因素, 在一定条件下大豆百粒重与大豆产量呈显著正相关^[3]。*GW2* 基因在多种植物中发挥调控籽粒性状的功能。目前, 已在水稻^[4-5]、玉米^[6-7]、大麦^[8]、小麦^[9-10]、葡

萄^[11]、拟南芥^[12-13] 中对 *GW2* 基因开展了一定的研究。Song 等^[4] 发现 *OsGW2* 是控制水稻粒宽和粒重的主效基因, 并利用图位克隆的方法成功分离该基因, 明确其位于水稻 2 号染色体上, 编码 RING 型蛋白, 具有 E3 泛素连接酶活性。该蛋白通过将其底物锚定到蛋白酶体进行降解, 从而负调节细胞的分裂, 进而导致穗壳宽度增加, 并且加速了籽粒灌浆的速率, 导致籽粒宽度、重量和产量增加。在小麦中利用同源克隆的方法成功克隆了 *OsGW2* 基因的同源基因, 在 61 ~ 103 位氨基酸处发现 1 个具有 E3 泛素连接酶活性的 RING 结构域, 其蛋白也能增加

小麦的粒宽和粒重^[9]。在拟南芥中也成功克隆了 *OsGW2* 的同源基因 *AtGW2*, 编码的蛋白质也含有 1 个 RING 结构域, 初步证明 *AtGW2* 基因可能通过某种机理负调控籽粒的大小^[12]。

大豆 *GW2* 基因还没有得到充分研究, 本实验室已在前期通过同源克隆的方法成功获得大豆 *GmGW2* 的 cDNA, 并进行了载体构建与遗传转化等初步研究^[14]。在此基础上, 本研究从大豆基因组中克隆出 *GmGW2* 基因的 DNA 序列, 并用生物信息学的方法对其进行较为详细的分析和预测, 旨在为后期的分子育种研究和挖掘 *GmGW2* 基因功能奠定理论基础, 进而为提高大豆产量的深入研究提供有益信息。

1 材料与方法

1.1 材料

供试大豆品种中黄 10 号由中国农业科学院作物科学研究所提供, 并在北京农学院实验田进行繁育。克隆载体 pCloneEZ 购自全式金公司。质粒提取试剂盒和 DNA 胶回收试剂盒购自美国 Omega 公司, 引物合成、抗生素、主要生化与分子生物学试剂均购自上海生工生物工程有限公司, 大肠杆菌菌株 DH5 α 由本实验室留存。

1.2 方法

1.2.1 大豆 *GmGW2* 基因的克隆 根据前期通过同源克隆得到的大豆中黄 10 号 *GmGW2* 基因的 cDNA 序列, 在 NCBI 上查找大豆基因组中同源性最高的序列。由于 *GmGW2* 的 DNA 序列较长, PCR 扩增比较困难且稳定性不好, 所以利用 Prime 5 软件设计 3 对特异性引物(表 1), 进行分段扩增。

表 1 *GmGW2* 特异性引物序列
Table 1 Specific primers of *GmGW2*

引物名称	引物序列(5' - 3')	片段长度
Name of primers	Sequence of primers(5' - 3')	Length of fragments
F1	TAAACCAGCACCTCAACC	3435
R1	AAGCCCTGAAACCCTACA	
F2	GCTATCTCATTTGGCTTTT	2566
R2	ATTCTTCTATTTCATTCG	
F3	AGCAGCAATGAATAGAAGAA	2803
R3	ATGGGTAATAAACTTGGGAG	

用 CTAB 法提取大豆中黄 10 号基因组 DNA。以提取的 DNA 为模板, 利用特异性引物扩增 *GmGW2* 基因序列, 反应体系(50 μ L): 5 \times Buffer 10 μ L、dNTP Mix 5 μ L、正反向引物各 1 μ L、模板 DNA 3 μ L(约 100 ng)、*FastPfu* DNA Ploymerase 1 μ L、ddH₂O 29 μ L。反应条件: 95 $^{\circ}$ C 预变性 5 min; 95 $^{\circ}$ C 变

性 30 s, 53 $^{\circ}$ C 退火 30 s, 72 $^{\circ}$ C 延伸 3.5 min, 35 个循环; 72 $^{\circ}$ C 延伸 10 min。将 PCR 产物经过 1% 的琼脂糖凝胶电泳检测, 得到目的片段, 将目的片段亚克隆到 pCloneEZ 载体上并测序。

1.2.2 *GmGW2* 基因的生物信息学分析

(1) 基因结构和系统进化树的构建

将测序获得的 3 段序列用 DNAMAN 软件进行序列分析与拼接后得到 *GmGW2* 的 DNA 序列全长。利用 GSDS 在线工具制作基因外显子 - 内含子结构示意图。将获得的 DNA 序列用 NCBI 网站上的在线软件 BLAST 进行序列相似性比对分析, 最终选择 14 种植物的同源序列, 运用 MEGA 5.0 软件构建 *GmGW2* 基因的系统发育树。

(2) 蛋白质理化性质和一级结构分析

利用 Expasy 网站的 ProtParamtool 对蛋白质进行理化性质和一级结构分析, 利用 Expasy 网站的在线工具 ProtScale 预测其蛋白质的亲水性与疏水性。

(3) 蛋白质二级结构、三级结构预测

用 SOPMA 在线软件对其编码蛋白的二级结构进行预测; 使用 SWISS-MODEL 在线软件预测蛋白质三级结构。

(4) 蛋白序列特征和功能分析

采用 TRMHMM server V 2.0 和 SignalP 4.1 分别进行跨膜域分析和预测信号肽, 利用 PSORT II Prediction 分析 *GmGW2* 蛋白的亚细胞定位; 使用 SWISS-MODEL 在线软件预测蛋白质三级结构; 采用 MOTIF search 和 Interpro 分析蛋白质的基序和结构域。

2 结果与分析

2.1 大豆 *GmGW2* 基因全长的克隆

1% 琼脂糖凝胶电泳检测显示提取的 DNA 条带清晰, 完整性好, 可进行下一步试验。通过 PCR 分段扩增获得 *GmGW2* 基因 DNA 片段一、二、三(图 1), 测序结果拼接出的 *GmGW2* 全基因序列与 NCBI 网站中的同源序列相似度为 98%。

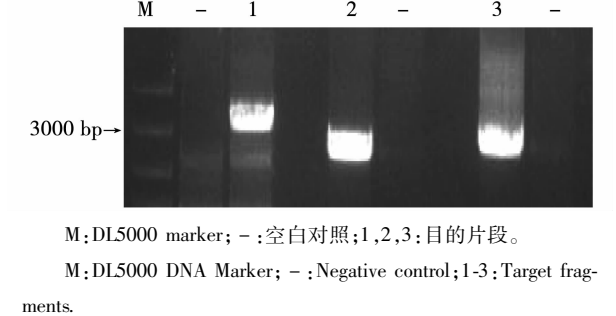


图 1 *GmGW2* 的分段扩增片段
Fig. 1 Segmented amplification of *GmGW2*

2.2 *GmGW2* 基因结构和系统进化树的构建

测序并拼接的结果显示:该基因 DNA 序列全长为8 467 bp,包含 8 个外显子和 7 个内含子。该基因为断裂基因,外显子和内含子的剪接不符合 GT-AG 剪接原则。DNAMAN 软件分析得出 *GmGW2* 基因序列中 A + T 碱基含量占 63.5% ,G + C 碱基含量占

36.5% (图 2)。

系统进化树分析结果显示:*GmGW2* 基因与木豆 (*Cajanus cajan*)、相思子(*Abrus precatorius*)和鹰嘴豆 (*Cicer arietinum*) 亲缘关系较近,与苜蓿 (*Medicago truncatula*) 关系较远(图 3)。

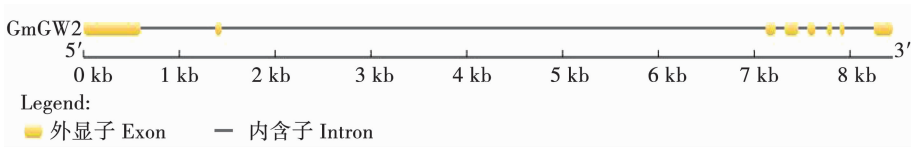


图 2 *GmGW2* 基因的结构
Fig.2 Structure of *GmGW2* Gene

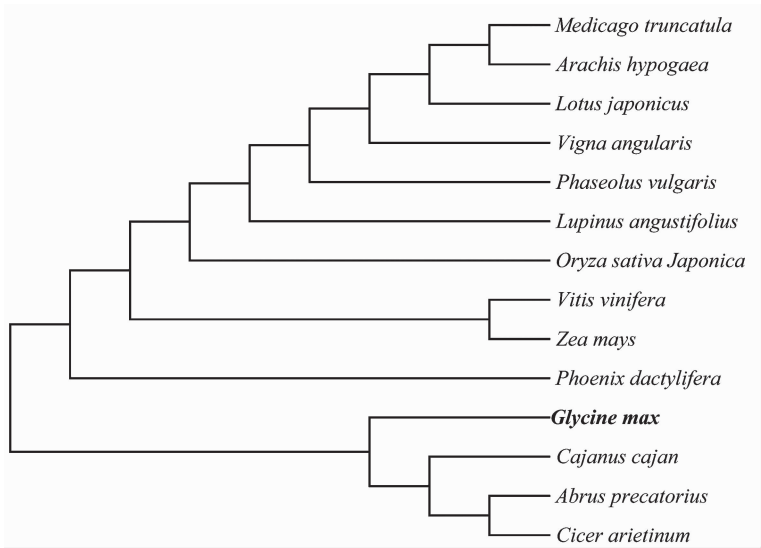


图 3 *GmGW2* 基因 DNA 序列的系统进化树
Fig.3 Phylogenetic tree of *GmGW2* gene DNA sequence

2.3 *GmGW2* 基因蛋白序列信息及结构分析

2.3.1 蛋白质理化性质和一级结构分析 通过 ExPASy 网站分析表明 *GmGW2* 基因编码蛋白的分子式为 C₂₀₂₅H₃₂₂₄N₅₉₄O₆₉₄S₂₉,总共6 566个原子,理论分子质量为 47.93 ku,等电点为 4.88,编码 431 个氨基酸,其中 46 个带正电荷氨基酸(Arg + Lys),66 个带负电荷氨基酸(Asp + Glu),不稳定参数为 70.05,为不稳定蛋白。氨基酸组成分析表明:丝氨酸含量最高(12.3%),其次是谷氨酸(9.5%),含量最低的是色氨酸(0.7%)。脂肪系数为 63.81,总平均疏水性为 -0.647。该蛋白预测为弱酸性蛋白。

蛋白质的亲水性对于维持蛋白质高级结构非常重要。该基因蛋白亲水/疏水曲线结果显示:其小于 0 的数值稍多于大于 0 的数值,且该蛋白的最高值为 4.5,在第 37 个氨基酸处,疏水性最强;最低值为 -4.5,在第 7、8、9、234、235、236 氨基酸处,亲水性最强(图 4)。因此,大豆 *GmGW2* 蛋白是一种亲水性蛋白。

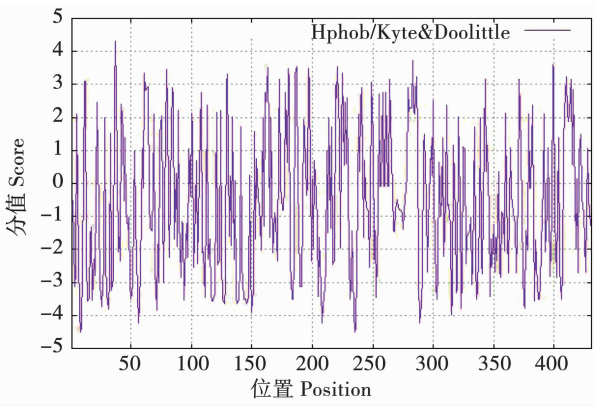


图 4 *GmGW2* 蛋白的亲水性/疏水性预测
Fig.4 Prediction of *GmGW2* protein hydrophilia/hydrophobic

2.3.2 蛋白质二级结构、三级结构预测 在线预测结果显示该蛋白二级结构、三级结构中不规则卷曲高达 60.56% ,次之 α -螺旋,所占比例为 26.22% , β -转角所占比例为 7.19% , β -折叠含量最少,仅占 6.03% 。表明 α -螺旋及不规则卷曲是其整体蛋

白质结构中的主要组成结构元件(图5)。通过SWISS-MODEL进行三级结构的同源建模(图6),找到该蛋白的同源模型 PDB:1jm7.1A,经过比对,该蛋白和1jm7.1A的氨基酸序列一致性达31.25%。

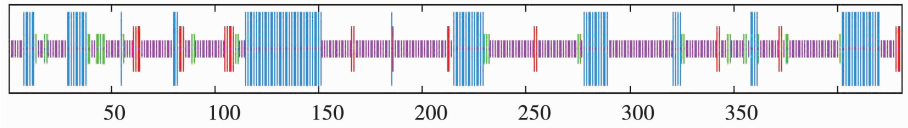


图5 GmGW2蛋白的二级结构预测

Fig. 5 Prediction of GmGW2 protein secondary structure



图6 GmGW2蛋白的三维结构图

Fig. 6 Prediction of GmGW2 protein swiss-model

2.4 蛋白质序列的特征和功能分析

2.4.1 蛋白质跨膜结构区分析、信号肽及亚细胞定位预测 跨膜结构区分析结果表明 GmGW2 基因编码蛋白质为非膜蛋白(图7)。信号肽预测结果表明 GmGW2 基因编码蛋白质的氨基酸残基信号肽最大分值没有超过 0.5(图8),说明该蛋白不存在信号肽,是一种非分泌蛋白,不能引导蛋白质的跨膜运输。

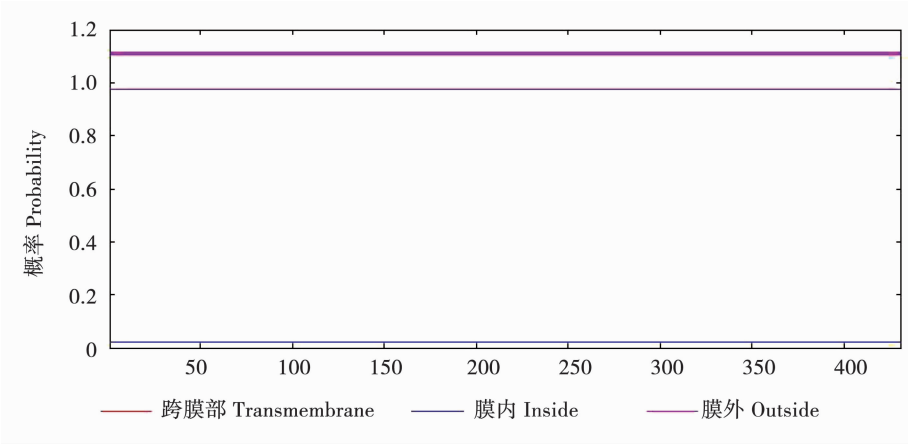


图7 GmGW2 编码蛋白的跨膜结构域分析

Fig. 7 Analysis on transmembrane domain of protein encoded by GmGW2

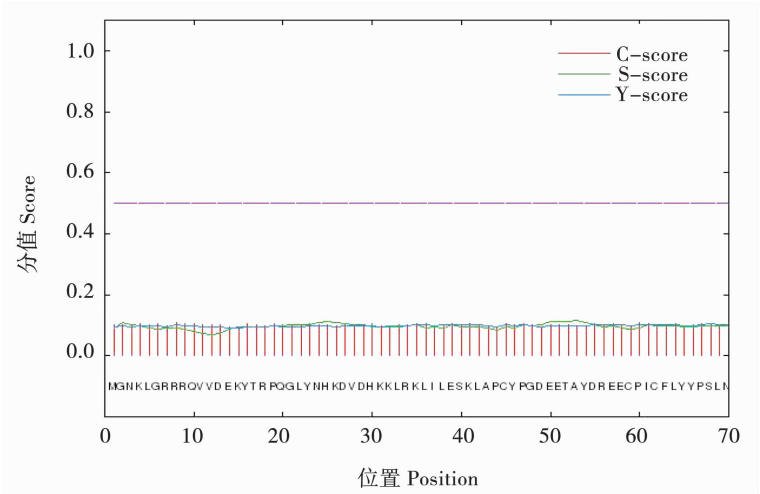


图8 GmGW2 编码蛋白信号肽预测

Fig. 8 Prediction on signal peptide of protein encoded by GmGW2

亚细胞定位分析结果显示:*GmGW2* 蛋白存在于细胞核、细胞质、细胞骨架和线粒体的概率分别为 43.5%、39.1%、8.7% 和 8.7%。推测该蛋白应该主要在细胞核或细胞质中发挥其生物学功能。

2.4.2 蛋白质基序和结构域分析 对 *GmGW2* 基因编码蛋白基序的分析结果显示:其含有 CDC73_N

基序(128 ~ 226)和 zf-C3HC4_3 基序(56 ~ 103)(图 9)。Interpro 结构域分析结果显示:该蛋白具有 Zinc finger 和 RING-type 的结构域(图 10)。这些特殊的基序和结构域常与蛋白质的某种生物学功能相关^[15]。

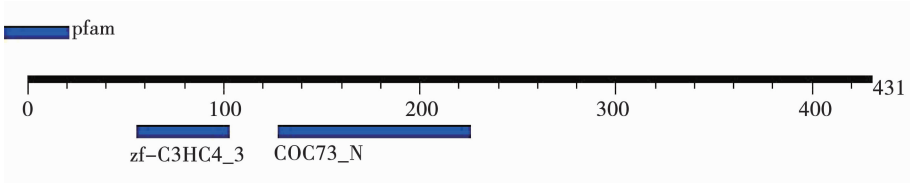


图 9 *GmGW2* 编码蛋白的 motif 分析
Fig. 9 Analysis on motif of protein encoded by *GmGW2*

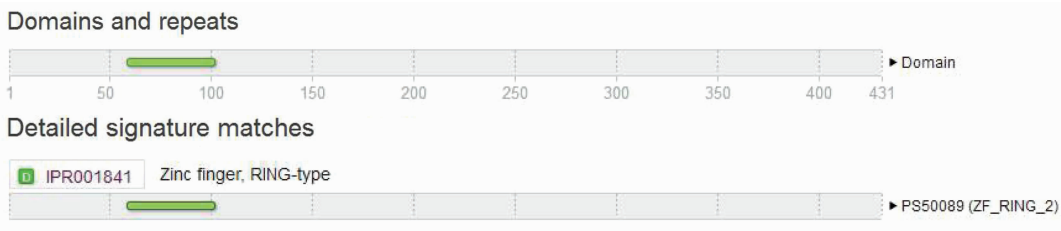


图 10 *GmGW2* 编码蛋白的结构域分析
Fig. 10 Analysis on structure domain of protein encoded by *GmGW2*

3 讨 论

生物信息学分析结果显示中黄 10 号大豆品种 *GmGW2* 基因外显子和内含子的剪接不符合 GT-AG 剪接原则,而在大豆品种 Williams 82 中该基因剪接机制则符合 GT-AG 剪接原则^[16],说明不同大豆品种的 *GmGW2* 基因在形成成熟的 mRNA 过程中其剪接机制存在差异,具体的原因和机制有待进一步的探索和研究。

GmGW2 基因的 DNA 序列系统进化树显示该基因与鹰嘴豆亲缘关系较近,然而前期利用该基因的氨基酸序列进行系统进化分析表明其与鹰嘴豆关系较远^[14]。说明该基因在 DNA 水平和氨基酸水平上的进化模式可能存在一定差异。

GmGW2 基因编码 431 个氨基酸,推测是一个定位于细胞核或细胞质,不含跨膜结构域,不含信号肽的不稳定亲水蛋白,而在水稻中 *OsGW2* 基因是定位于细胞质中,可以使水稻籽粒变小的负调控基因^[4,9]。大豆 *GmGW2* 的亚细胞定位预测结果与水稻 *OsGW2* 的亚细胞定位结果相一致,这对研究大豆 *GmGW2* 基因的功能提供了一定的参考。

Interpro 结构域分析表明 *GmGW2* 蛋白具有 RING 型结构域。大量研究表明,RING 型锌指蛋白广泛参与植物的生长发育、胁迫与信号转导等过

程^[17]。据报道,水稻中的 *OsGW2* 基因编码 1 个具有 E3 泛素连接酶活性的 RING 型蛋白(C5HC2),这种酶蛋白主要控制水稻谷粒的粒宽和粒重,同时参与调控植物逆境胁迫信号途径^[4]。因此推测 *GmGW2* 蛋白可能也具有类似的功能。

4 结 论

本文从大豆基因组中克隆出了 *GmGW2* 基因的全长序列,并在 DNA 水平和蛋白质水平上进行生物信息学分析。结果显示该基因为断裂基因,含 8 个外显子和 7 个内含子,其剪接不符合 GT-AG 剪接原则;*GmGW2* 基因的 DNA 序列系统进化树分析结果表明该基因与木豆、相思子、鹰嘴豆亲缘关系较近,与花生、苜蓿亲缘关系较远;*GmGW2* 蛋白分析表明该蛋白是一个定位于细胞核或细胞质中,不含跨膜结构域及信号肽,具有 Zinc finger 和 RING-type 结构域的不稳定亲水蛋白。根据 *GmGW2* 蛋白的结构域特征推测该基因可能具有调控大豆籽粒性状的功能。本研究结果为后期进一步挖掘和验证 *GmGW2* 基因的功能奠定了理论基础,同时也为培育出高产的大豆品种提供了新思路。

参考文献

[1] 余永亮,梁慧珍,王树峰,等. 中国转基因大豆的研究进展及

其产业化[J]. 大豆科学, 2010, 29(1): 143-150. (Yu Y L, Liang H Z, Wang S F, et al. Research progress and commercialization on transgenic soybean in China [J]. Soybean Science, 2010, 29(1): 143-150.)

[2] 查霆, 钟宣伯, 周启政, 等. 我国大豆产业发展现状及振兴策略[J]. 大豆科学, 2018, 37(3): 135-140. (Zha T, Zhong X B, Zhou Q Z, et al. Development status of China's soybean industry and strategies of revitalizing[J]. Soybean Science, 2018, 37(3):135-140.)

[3] 邱丽娟, 马岩松, 刘章雄, 等. 群体构成方式对大豆百粒重全基因组选择预测准确度的影响[J]. 作物学报, 2018, 44(1): 43-52. (Qiu L J, Ma Y S, Liu Z X, et al. Effect of population structure on prediction accuracy of soybean 100-seed weight by genomic selection [J]. Acta Agronomica Sinica, 2018, 44(1): 43-52.)

[4] Song X J, Huang W, Shi M, et al. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase[J]. Nature Genetics, 2007, 39(5):623-630.

[5] Zhang Y D, Zheng J, Liang Y L, et al. Functional marker development and effect analysis of grain size gene *GW2* in extreme grain size germplasm in rice[J]. Rice Science, 2015, 22(2):65-70.

[6] Q, Lin L, Yang X H, et al. Relationship, evolutionary fate and function of two maize co-orthologs of rice *GW2* associated with kernel size and weight [J]. BMC Plant Biology, 2010, 10(1):143.

[7] 孔曼丽, 李彩云, 孙清鹏, 等. 玉米 E3 泛素连接酶类基因 *ZmGW2-1* 的克隆及表达分析[J]. 安徽农业大学学报, 2014, 41(6):1004-1011. (Kong M L, Li C Y, Sun Q P, et al. Isolation and expression analysis of the E3 ubiquitin ligase encoding gene *ZmGW2-1* in maize[J]. Journal of Anhui Agriculture University, 2014, 41(6): 1004-1011.)

[8] 张雨良, 王希东, 张桦, 等. 大麦产量相关基因 *HvYrg1* 的克隆及植物 RNA 干扰载体的构建[J]. 植物研究, 2009, 29(4):445-452. (Zhang Y L, Wang X D, Zhang H, et al. Cloning of hordeum vulgare yield related gene *HvYrg1* and construction of its plant RNA interfere vector[J]. Bulletin of Botanical Research, 2009, 29(4):445-452.)

[9] Su Z, Hao C, Wang L, et al. Identification and development of a functional marker of *TaGW2* associated with grain weight in bread wheat [J]. Theoretical and Applied Genetics, 2011, 122(1): 211-223.

[10] James S, Peter S, Jemima B, et al. A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains[J]. Theoretical Applied Genetics, 2016, 129(6):1099-1112.

[11] 宋新新, 袁月, 孙海龙, 等. 葡萄 *VvGW2* 基因的克隆及表达[J]. 江苏农业学报, 2015, 31(2):434-440. (Song X X, Yuan Y, Sun H L, et al. Cloning and expression of *VvGW2* gene in grape[J]. Jiangsu Journal of Agricultural Sciences, 2015, 31(2):434-440.)

[12] 李会勇, 曹言勇, 路运才, 等. 拟南芥中 RING 型 E3 泛素连接酶基因 *AtGW2* 的克隆和功能分析[J]. 植物遗传资源学报, 2011, 12(3): 448-454. (Li H Y, Cao Y Y, Lu Y C, et al. Cloning and functional analysis of *Arabidopsis thaliana AtGW2*, a RING-type E3 ubiquitin ligase protein[J]. Journal of Plant Genetic Resources, 2011, 12(3): 448-454.)

[13] Xia T, Li N, Dumenil J, et al. The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in *Arabidopsis*[J]. Plant Cell, 2013, 25(9):3347-3359.

[14] 熊鹤雯, 吴志慧, 段思宇, 等. 大豆基因 *GmGW2* 的克隆与转化[J]. 生物技术通报, 2015(8): 82-87. (Xiong H W, Wu Z H, et al. Cloning and transformation of gene *GmGW2* in soybean [J]. Biotechnology Bulletin, 2015(8): 82-87.)

[15] 王举. 生物信息学基础及应用[M]. 北京: 清华大学出版社, 2014, 126-130. (Wang J. Bioinformatics basics and applications [M]. Bei Jing: Tsinghua University Press, 2014:126-130.)

[16] Jeremy S, Steven B, Jessica S, et al. Genome sequence of the palaeopolyploid soybean [J]. Nature, 2010, 463 (7278): 178-183.

[17] 陆瑜. 植物中 RING 型锌指蛋白研究进展[J]. 安徽农学通报, 2017(8):42-45. (Lu Y. Research progress of ring-type zinc finger protein in plants[J]. Anhui Agricultural Science Bulletin, 2017(8):42-45.)