

南农 87C-38 大豆优质 11S 球蛋白 Gy7 基因克隆及序列分析

顾婷玉, 李建粤, 米东, 肖刚

(上海师范大学生命与环境科学学院, 上海 200234)

摘要 大豆种子不仅富含蛋白质, 而且赖氨酸含量较高, 利用大豆高赖氨酸基因能够弥补谷类作物种子赖氨酸含量的不足。运用现代分子生物学技术, 从高蛋白大豆南农 87C-38 中克隆 11S 球蛋白 Gy7 全基因及 cDNA 序列。经生物公司测序后, 利用 vector NTI 软件将所克隆的 Gy7 全基因及 cDNA 序列与 Genbank (AF319776 和 AF319777) 报道的序列进行比对分析, 同源性分别为 99% 和 99.6%。序列比对结果显示, Gy7 基因 cDNA 与 Genbank 报道的序列之间所有的核苷酸差异都位于第 3 外显子。由此推测, 第 1、第 2 和第 4 外显子编码的氨基酸对于 G7 多肽形成特定高级结构可能具有非常重要的作用, 它们在进化过程中受到的选择压力可能比第 3 外显子更强。根据遗传密码表推测, 克隆的 Gy7 与 Genbank 报道的 cDNA 序列所编码的多肽, 两者存在 2 个氨基酸组成的差异。大豆球蛋白 Gy7 优质基因的获得为今后利用转基因技术改良水稻、小麦等谷类作物的营养品质奠定了基础。

关键词 大豆; 11S 球蛋白 Gy7 全基因; 11S 球蛋白 Gy7 cDNA 序列; 克隆; 序列分析

中图分类号 S565.1 **文献标识码** A **文章编号** 1000-9841(2007)06-0840-07

CLONING AND SEQUENCE ANALYSES OF 11S GLYCININ Gy7 GENE FROM SOYBEAN NANNONG 87C-38

GU Ting-yu, LI Jian-yue, MI Dong, XIAO Gang

(College of Life and Environment Sciences, Shanghai Normal University, Shanghai 200234)

Abstract Rice and wheat are main crops for human consumption, but the seeds in those crops generally lack lysine. Soybean has an abundant storage of protein and lysine, which could help to make up for the lack of lysine in rice and wheat. 11S Glycinin Gy7 gene and its cDNA of soybean Nannong 87C-38 were cloned, independently, with the help of the molecular biotechnology. After being compared the two sequences with Genbank (AF319776 and AF319777) by vector NTI software, their identities were 99% and 99.6%. Interestingly, most differences of the nucleotide acid between Gy7 cDNA and Genbank AF319777 were located in the third exon. Thus we presume that the amino acid coded by the first, second and fourth exons are more stable and they play a more important role in forming specific structures of the G7 polypeptide than the third exon. According to the genetic code, we presumed the amino acid

收稿日期: 2007-06-01

基金项目: 上海市科学技术委员会资助项目 (063919141)

作者简介: 顾婷玉 (1982-), 女, 硕士研究生, 研究方向为植物分子遗传学。

通讯作者: 李建粤, 副教授, 研究生导师。E-mail: lijianyue01@yahoo.com.cn

coded by *Gy7* cDNA. There were two amino acid differences from Genbank AF319777. This superior gene cloned in this research could lay the foundation for improving nutritional quality of the crops such as rice and wheat.

Key words Soybean; 11S glycinin *Gy7* gene; 11S glycinin *Gy7* cDNA; Clone; Sequence analyses

在水稻、小麦等谷物种子中,不仅贮藏蛋白含量较少^[1],而且赖氨酸含量也比较低^[2],是限制谷类作物种子营养品质的第一限制性氨基酸。传统育种难以有效地提高作物必需氨基酸含量,但采用转基因技术有可能实现快速有效地改良谷类作物种子营养品质的目标,然而具备优质的贮藏蛋白基因是实现这一目标的首要条件。大豆种子富含蛋白质,其含量约为种子总重量的40%。而且,在大豆种子中赖氨酸含量较高,是谷物赖氨酸含量的10倍^[3],完全能够弥补谷类作物种子赖氨酸含量的不足。

在大豆种子蛋白质中90%是盐溶性球蛋白^[4]。按照在一定pH和盐离子浓度环境中沉降系数不同,盐溶性球蛋白又可分为2S、7S、11S和15S4种组分。其中7S和11S球蛋白是两种主要成分,分别被称为豌豆球蛋白(Vicilin)和豆球蛋白(Legumin)^[5]。大豆的豆球蛋白主要为大豆球蛋白(glycinin),约占大豆成熟种子总蛋白含量的50%^[6]。

迄今为止已报道了6种有功能的大豆球蛋白基因,它们分别是*Gy1*~*Gy5*以及*Gy7*基因^[7-9]。李建粤等^[10]分析比较了在Genbank中报道的*Gy1*~*Gy5*和*Gy7*6种有功能大豆球基因编码的氨基酸组成后提出,大豆球蛋白*Gy7*基因在6种大豆球蛋白中赖氨酸含量最高,更适合作为改良水稻等禾谷类作物营养品质的目的基因。

以高蛋白大豆南农87C-38为材料,利用分子生物学常规操作技术,克隆了优质的大豆球蛋白*Gy7*全基因及cDNA序列,并将克隆的*Gy7*基因与Genbank中报道的序列进行比对。为利用基因工程技术改良水稻、小麦等谷类作物种子营养品质奠定基础。

1 材料与方法

1.1 植物材料

高蛋白大豆(*Glycine max* (L.) Merri.)南农87C-38,由南京农业大学大豆研究所邱家驯研究员提供。

1.2 试剂、质粒和菌株

PCR扩增试剂、PCR产物回收试剂盒、反转录试剂盒、T₄DNA连接酶及pUCm-T载体等购自上海生工生物工程有限公司。pMD18-T载体、保真性能较好的LA-Taq DNA聚合酶、各种限制性内切酶等由Takara公司购得。大肠杆菌菌株为DH5 α 。

1.3 方法

1.3.1 大豆球蛋白*Gy7*总基因的获得 根据Genbank (AF319776)公布的大豆球蛋白*Gy7*基因序列在翻译起始点ATG上游28bp处和终止密码子TAA后92bp处分别设计特异PCR引物,在上游引物一端引入XbaI位点,在下游引物一端引入XhoI位点。上游引物序列为:5'-G TCTAGA GCTATCCCTCG-CATCTGTTTCAG-3'(下划线碱基表示XbaI酶切位点),下游引物序列为:5'-G CTCTGA GGTAGCAT-TCGATCCATAATGTG-3'(下划线碱基表示XhoI酶切位点)。引物由上海生工生物工程有限公司合成。以南农87C-38大豆总DNA为模板,利用PCR扩增获得大豆球蛋白*Gy7*总基因序列。PCR反应程序:94℃预变性5 min;94℃变性1 min,56℃退火2 min,72℃延伸2 min,40个循环,72℃保温10 min。PCR产物经1%琼脂糖凝胶电泳检测,根据分子量大小初步确定是所要扩增的目的片段后,按照上海生工生物工程有限公司回收试剂盒操作指南回收目的片段。按照试剂盒操作说明将回收的目的片段克隆到pUCm-T载体上。质粒DNA的提取、纯化、PCR、酶切鉴定和转化等操作参照分子克隆实验指南进行^[11]。选取载有目的片段的阳性pUCm-T克隆载体送上海基康生物公司进行测序分析。

1.3.2 大豆球蛋白*Gy7*基因cDNA的获得 摘取未成熟的南农87C-38大豆豆荚,从大豆子叶中提取总RNA,利用反转录试剂盒获得cDNA的第一链。根据Genbank AF319777公布的大豆球蛋白*Gy7* mRNA序列在翻译起始点ATG上游4bp处和终止密码子TAA处分别设计特异PCR引物,同样在上游引物一端引入XbaI位点,在下游引物一端引入XhoI位点。上游引物序列为:5'-G TCTAGACAC-

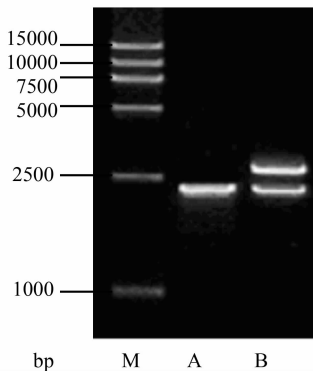
CATGTTTAACCATTTCTGCGC-3' (下划线碱基表示 XbaI 酶切位点), 下游引物序列为: 5'-G CTGAG TTACATGGTGACAATGAGGGGAT-3' (下划线碱基表示 XhoI 酶切位点)。以 cDNA 的第一链为模板, PCR 扩增获得大豆球蛋白 Gy7 基因的 cDNA 序列。

PCR 扩增的参数为: 预变性 94℃ 3 min; 变性 94℃ 1 min, 退火 60℃ 1min, 延伸 72℃ 2 min, 40 循环; 最后保温 72℃ 10 min。电泳回收目的片段重组到 pMD18-T 载体上, 转化大肠杆菌感受态细胞, PCR、酶切鉴定筛选出阳性克隆后送上海基康生物公司进行测序分析。

2 结果与分析

2.1 Gy7 全基因序列与 pUCm-T 载体的连接鉴定

Gy7 全基因与 pUCm-T 载体连接后, 转化大肠杆菌, 用 Amp、X-gal、IPTG 筛选培养, 经 12 ~ 16 h 培养后长出菌落, 挑选白色菌落, 提取质粒, 进行 PCR 和 Xba I、Xho I 的双酶切鉴定, 结果如见 1。PCR 和双酶切产物均有大小约为 2200bp 的条带, 其大小与预期 Gy7 全基因片段条带大小一致。双酶切产物的另一条带大约为 2700bp, 与 pUCm-T 载体的大小一致, 说明 Gy7 全基因序列成功构建入 pUCm-T 载体。



M: 标准 DNA; A: PCR 产物;
B: Xba I、Xho I 双酶切产物

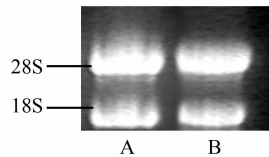
M: DNA Marker; A: A product of PCR of Gy7;
B: Products of double digestion of positive clone

图 1 重组阳性克隆 PCR 及双酶切检测

Fig. 1 Identification of positive clone by PCR and digestion by double enzyme

2.2 Gy7 基因 cDNA 制备及 cDNA 序列与 pUCm-T 载体的连接鉴定

从未成熟大豆种子中提取的总 RNA 经甲醛变性胶的电泳后, 呈现出 2 条清晰可区分的条带, 分别为 28S 和 18S (图 2), 两者的亮度比约为 2: 1, 表明 RNA 基本无降解。



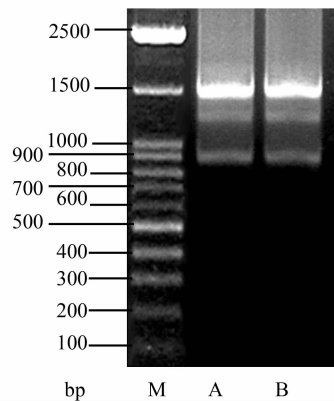
A - B: 大豆总 RNA

A - B: Total RNA in soybean

图 2 大豆总 RNA 的电泳检测

Fig. 2 Identification of total RNA in soybean

以逆转录产物 cDNA 第一条链为模板, 用人工合成的引物进行扩增反应, 电泳检测有 3 条带, 结果如图 3 所示。该现象与 Beilinson 等^[9]报道结果一致。Beilinson 等认为, 其中 1600bp 的条带即为 Gy7 基因 cDNA 链, 第二条条带约占所有扩增产物的 30%, 该片段缺少了 Gy7 基因第三外显子。第三条最小的条带则是未知区域的一段序列, 它与 Gy7 基因无关。



M: 标准 DNA; A - B: RT-PCR 扩增产物

M: DNA Marker; A - B: RT-PCR products

图 3 RT-PCR 产物琼脂糖电泳检测

Fig. 3 Identification of RT-PCR products

Gy7 cDNA 与 pMD18-T 载体连接后也转化大肠杆菌, 经过 37℃ 培养约 12 ~ 16 h 后挑选白色菌落, 提取质粒, 进行 PCR 和 Xba I、Xho I 的双酶切检测 (如图 4)。双酶切和 PCR 产物均有大小约为 1600bp 的条带, 其大小与预期 Gy7 基因 cDNA 一致。双酶切的另一条带大约为 2700bp, 与 pMD18 -

T 载体的大小一致,由此表明 *Gy7* cDNA 片段已成功构建入 pMD18 - T 载体中。

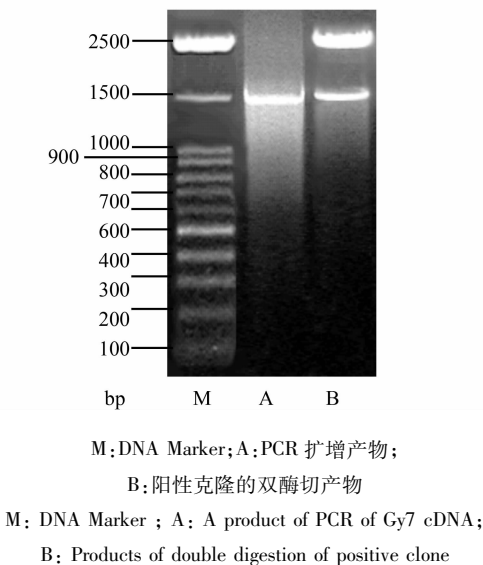


图 4 阳性克隆的 PCR 及双酶切检测

Fig. 4 Identification of positive clone by PCR and digestion by double enzyme

2.3 *Gy7* 全基因及 cDNA 测序结果分别与 Genbank 对应序列比对

将连接上 *Gy7* 全基因及 cDNA 片段的 T 载体送上海基康生物公司进行测序。测序结果显示,克隆的 *Gy7* 全基因从翻译起始位点 ATG 至终止密码子 TAA 核苷酸全长为 2171bp,对应的 cDNA 核苷酸总长为 1611bp

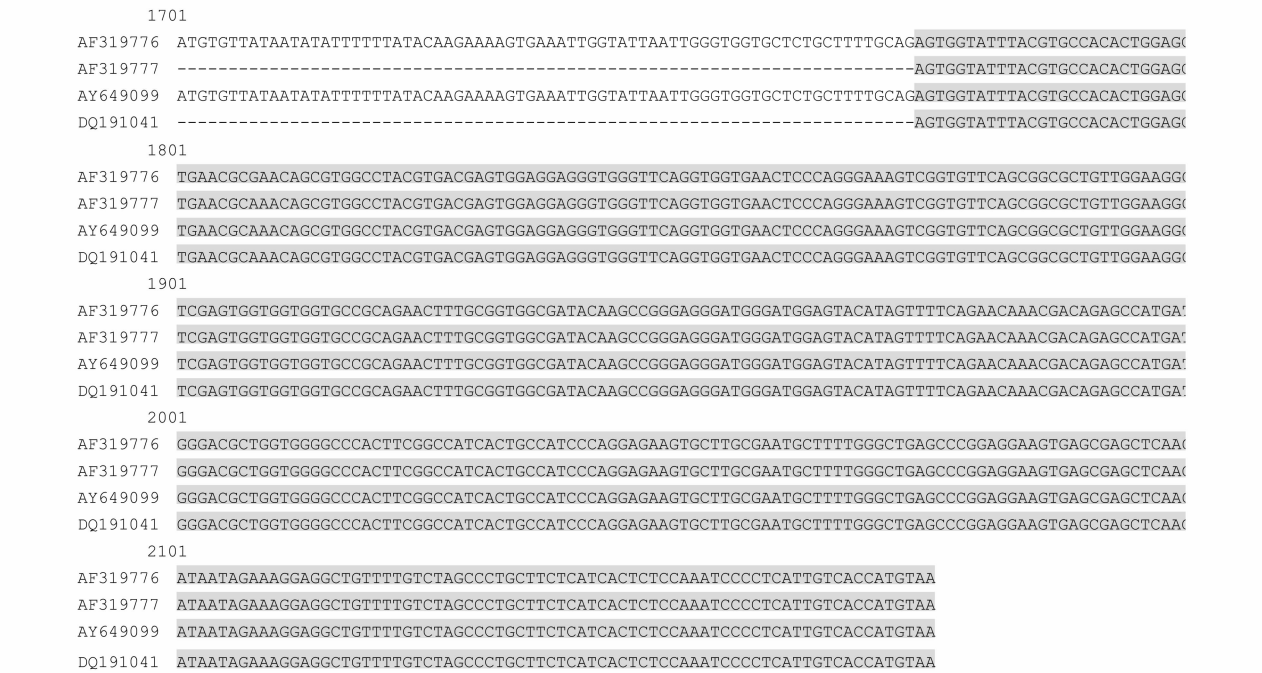
	1
AF319776	ATGTTTAACCATCTGCGCTCCATTATTATTTTCTCCTCTTCTTCACGTGCACGTGCTTAGCACGACAGCAATGCCAGTTCAAGCAGGAGTGCCAACT
AF319777	ATGTTTAACCATCTGCGCTCCATTATTATTTTCTCCTCTTCTTCACGTGCACGTGCTTAGCACGACAGCAATGCCAGTTCAAGCAGGAGTGCCAACT
AY649099	ATGTTTAACCATCTGCGCTCCATTATTATTTTCTCCTCTTCTTCACGTGCACGTGCTTAGCACGACAGCAATGCCAGTTCAAGCAGGAGTGCCAACT
DQ191041	ATGTTTAACCATCTGCGCTCCATTATTATTTTCTCCTCTTCTTCACGTGCACGTGCTTAGCACGACAGCAATGCCAGTTCAAGCAGGAGTGCCAACT
	101
AF319776	ATACCATCCATGCACTGAAACCTGACAACCTCATCGAATCCCAAGGTGGTGTACAGAGACATGGAACGCTAGCCACCCTGAGCTATGTTGCGCTGGC
AF319777	ATACCATCCATGCACTGAAACCTGACAACCTCATCGAATCCCAAGGTGGTGTACAGAGACATGGAACGCTAGCCACCCTGAGCTATGTTGCGCTGGC
AY649099	ATACCATCCATGCACTGAAACCTGACAACCTCATCGAATCCCAAGGTGGTGTACAGAGACATGGAACGCTAGCCACCCTGAGCTATGTTGCGCTGGC
DQ191041	ATACCATCCATGCACTGAAACCTGACAACCTCATCGAATCCCAAGGTGGTGTACAGAGACATGGAACGCTAGCCACCCTGAGCTATGTTGCGCTGGC
	201
AF319776	CGCCTTTATCAAGCGCACCATCAACCCCAATGGCCTTCACTTGCCATCCTACGTTAATTACCCGGAACCTCCATTTTCGTACTCCAAGGTATCCAAACA
AF319777	CGCCTTTATCAAGCGCACCATCAACCCCAATGGCCTTCACTTGCCATCCTACGTTAATTACCCGGAACCTCCATTTTCGTACTCCAAGGT-----
AY649099	CGCCTTTATCAAGCGCACCATCAACCCCAATGGCCTTCACTTGCCATCCTACGTTAATTACCCGGAACCTCCATTTTCGTACTCCAAGGTATCCAAACA
DQ191041	CGCCTTTATCAAGCGCACCATCAACCCCAATGGCCTTCACTTGCCATCCTACGTTAATTACCCGGAACCTCCATTTTCGTACTCCAAGGT-----
	301
AF319776	TCATATTCATTCTCCTCATCAATTAATCAATTATCTTATTCAAGTTAATTAATCTAGTAAATTTGTATACTAACTAATKATKTTTATTATTTTAT
AF319777	-----
AY649099	TCATATTCATTCTCCTCATCAATTAATCAATTATCTTATTCAAGTTAATTAATCTAGTAAATTTGTATACTAACTAATATTTTATTATTTTAT
DQ191041	-----

利用 Vector NTI 软件分别将本试验克隆的 *Gy7* 全基因和 cDNA 从起始密码子至终止密码子序列分别与 Genbank 报道 *Gy7* 全基因 (AF319776) 和 cDNA 序列 (AF319777) 进行同源性比较,结果如图 5 所示。克隆的 *Gy7* 基因也含有 3 个内含子,4 个外显子(图 5 阴影显示)。该结果与已报道的其他 5 种有功能的大豆球蛋白基因的结构也是一致的^[12]。南农 87C-38 *Gy7* 基因与 Genbank 报道的 *Gy7* 基因编码序列有 6 个碱基不同(图 5 方框显示),同源性为 99.6%。在内含子中,南农 87C-38 *Gy7* 基因与 Genbank 报道的 *Gy7* 基因约有 18 个碱基差异,有些位点是碱基的改变,有些是碱基的插入或碱基的缺失。两个 *Gy7* 全基因序列同源性约为 99.0%。

比对结果表明,克隆所获得的 DNA 片段的确定为 *Gy7* 全基因及其 cDNA 片段,并且与 GenBank 中 *Gy7* 基因序列存在品种差异性。这两个序列已提交 Genbank,注册号分别为 AY649099 和 DQ191041。

克隆的 *Gy7* 基因与 GenBank 报道的编码多肽产物长度相同,都含有 536 个氨基酸残基。根据遗传密码表,推测两种大豆品种 *Gy7* 基因编码的 G7 多肽氨基酸组成。克隆的 *Gy7* 基因 cDNA 与 Genbank 报道 cDNA 序列有 6 个核苷酸不同,其中有 4 个位点核苷酸差异没有产生氨基酸种类的变化,另 2 个核苷酸差异有氨基酸组成变化,主要表现是酸性氨基酸与中性氨基酸之间的改变(表 1)。

	401	
AF319776	ATAAATATTAAGCTTTTTTAATTATCCAACCARTCTTATATATTTTGTATATTTAATTATGAGTTGACAGGTGAGGGAGTGTGGGAATTGTAATTC	
AF319777	-----GAGGGAGTGTGGGAATTGTAATTC	
AY649099	ATAAATATTAAGCTTTTTTAATTATCCAACCAATCTTATATATTTGGTATATTTAATTGATGAGTTGACAGGTGAGGGAGTGTGGGAATTGTAATTC	
DQ191041	-----GAGGGAGTGTGGGAATTGTAATTC	
	501	
AF319776	GGTTGTGACGAAACTTTTGAAGAGCCACAACGGGAGAGGGAACATGATCGCCACCAGAAGGTCGTTACCTGAAGCAGGGTGACATATTCGCAGTTCC	
AF319777	GGTTGTGACGAAACTTTTGAAGAGCCACAACGGGAGAGGGAACATGATCGCCACCAGAAGGTCGTTACCTGAAGCAGGGTGACATATTCGCAGTTCC	
AY649099	GGTTGTGACGAAACTTTTGAAGAGCCACAACGGGAGAGGGAACATGATCGCCACCAGAAGGTCGTTACCTGAAGCAGGGTGACATATTCGCAGTTCC	
DQ191041	GGTTGTGACGAAACTTTTGAAGAGCCACAACGGGAGAGGGAACATGATCGCCACCAGAAGGTCGTTACCTGAAGCAGGGTGACATATTCGCAGTTCC	
	601	
AF319776	CTGGAATTCCTTACTGGACCTACAACCTACGCCAATGTTTCTCTTGTGTAATTACCCCTCCTTGACACTGCCAATTCGAAAAACAGCTTGATCGTGTCT	
AF319777	CTGGAATTCCTTACTGGACCTACAACCTACGCCAATGTTTCTCTTGTGTAATTACCCCTCCTTGACACTGCCAATTCGAAAAACAGCTTGATCGTGTCT	
AY649099	CTGGAATTCCTTACTGGACCTACAACCTACGCCAATGTTTCTCTTGTGTAATTACCCCTCCTTGACACTGCCAATTCGAAAAACAGCTTGATCGTGTCT	
DQ191041	CTGGAATTCCTTACTGGACCTACAACCTACGCCAATGTTTCTCTTGTGTAATTACCCCTCCTTGACACTGCCAATTCGAAAAACAGCTTGATCGTGTCT	
	701	
AF319776	CAGAGTAAGAAATATTCGGCTTCACATGTATAGTATATACTAGTTCATCATATATATTTGGTGATGATGTTAAGGATGTTTTTTTTTTTTTTTCT	
AF319777	CAGA-----	
AY649099	CAGAGTAAGAAATATTCGGCTTCACATGTATAGTATATACTAGTTCATCATATATATTTGGTGATGATGTTAAGGATGTTTTTTTTTAAATTTT---CT	
DQ191041	CAGA-----	
	801	
AF319776	TTTCATTCCCTTGACATTATTAATATGGTGGTATATATTAGAGATTCTATCTTGCTGGGAACCCAAAAGAAGAGCACCCCTGTGGACGCAAGCAAGAA	
AF319777	-----AGATTCTATCTTGCTGGGAACCCAAAAGAAGAGCACCCCTGTGGACGCAAGCAAGAA	
AY649099	TTTCATTCCCTTGACATTATTAATATGGTGGTATATATTAGAGATTCTATCTTGCTGGGAACCCAAAAGAAGAGCACCCCTGTGGACGCAAGCAAGAA	
DQ191041	-----AGATTCTATCTTGCTGGGAACCCAAAAGAAGAGCACCCCTGTGGACGCAAGCAAGAA	
	901	
AF319776	AGGTAACAATATAAACATGTTCCGGTGGTTTCGACCCCGCGTTCTTAGCAGAAGCATCGAACGTGAAGGTGGGGATAACAAAGAAGCTTCAGTCACACA	
AF319777	AGGTAACAATATAAACATGTTCCGGTGGTTTCGACCCCGCGTTCTTAGCAGAAGCATCGAACGTGAAGGTGGGGATAACAAAGAAGCTTCAGTCACACA	
AY649099	AGGTAACAATATAAACATGTTCCGGTGGTTTCGACCCCGCGTTCTTAGCAGAAGCATCGAACGTGAAGGTGGGGATAACAAAGAAGCTTCAGTCACACA	
DQ191041	AGGTAACAATATAAACATGTTCCGGTGGTTTCGACCCCGCGTTCTTAGCAGAAGCATCGAACGTGAAGGTGGGGATAACAAAGAAGCTTCAGTCACACA	
	1001	
AF319776	GGTGACCAAAATCATAAAAGTGAAAAAGGCTCTTAGCATTATCAGGCCACCCCTTGGAAACACGAAGTTAGAGAAGCAGAAGTAGAAGAAAAACCAAGAC	
AF319777	GGTGACCAAAATCATAAAAGTGAAAAAGGCTCTTAGCATTATCAGGCCACCCCTTGGAAACACGAAGTTAGAGAAGCAGAAGTAGAAGAAAAACCAAGAC	
AY649099	GGTGACCAAAATCATAAAAGTGAAAAAGGCTCTTAGCATTATCAGGCCACCCCTTGGAAACACGAAGTTAGAGAAGCAGAAGTAGAAGAAAAACCAAGAC	
DQ191041	GGTGACCAAAATCATAAAAGTGAAAAAGGCTCTTAGCATTATCAGGCCACCCCTTGGAAACACGAAGTTAGAGAAGCAGAAGTAGAAGAAAAACCAAGAC	
	1101	
AF319776	GAGAACACTGTGAATGCCAAAAAGAAAGAAAAACAAAGAAGGAGAAGGAGAGGGAAGAGGTTGGAAGAGGTGGTACAAGAGAAAGAAATAAGAAAGCGCAAGCACCAC	
AF319777	GAGAACACTGTGAATGCCAAAAAGAAAGAAAAACAAAGAAGGAGAAGGAGAGGGAAGAGGTTGGAAGAGGTGGTACAAGAGAAAGAAATAAGAAAGCGCAAGCACCAC	
AY649099	GAGAACACTGTGAATGCCAAAAAGAAAGAAAAACAAAGAAGGAGAAGGAGAGGGAAGAGGTTGGAAGAGGTGGTACAAGAGAAAGAAATAAGAAAGCGCAAGCACCAC	
DQ191041	GAGAACACTGTGAATGCCAAAAAGAAAGAAAAACAAAGAAGGAGAAGGAGAGGGAAGAGGTTGGAAGAGGTGGTACAAGAGAAAGAAATAAGAAAGCGCAAGCACCAC	
	1201	
AF319776	AGGAGAACACGAGGGATGTGGTGAATGCGGAATATAAAGAAGAGGAAGAGCAAAAGTAGAAGCCGAGAACGCGCGAGTGGCATGAACATAAAGGACAAC	
AF319777	AGGAGAACACGAGGGATGTGGTGAATGCGGAATATAAAGAAGAGGAAGAGCAAAAGTAGAAGCCGAGAACGCGCGAGTGGCATGAACATAAAGGACAAC	
AY649099	AGGAGAACACGAGGGATGTGGTGAATGCGGAATATAAAGAAGAGGAAGAGCAAAAGTAGAAGCCGAGAACGCGCGAGTGGCATGAACATAAAGGACAAC	
DQ191041	AGGAGAACACGAGGGATGTGGTGAATGCGGAATATAAAGAAGAGGAAGAGCAAAAGTAGAAGCCGAGAACGCGCGAGTGGCATGAACATAAAGGACAAC	
	1301	
AF319776	CATGGAAAAGAAAAAGGAAGAGAGATATAAGGAAGGTGGTGAAGGGAGAGGCGGTAGCAATGTGTTGGAAGAAATCTTGTGCATTGGAAGCTGCA	
AF319777	CATGGAAAAGAAAAAGGAAGAGAGATATAAGGAAGGTGGTGAAGGGAGAGGCGGTAGCAATGTGTTGGAAGAAATCTTGTGCATTGGAAGCTGCA	
AY649099	CATGGAAAAGAAAAAGGAAGAGAGATATAAGGAAGGTGGTGAAGGGAGAGGCGGTAGCAATGTGTTGGAAGAAATCTTGTGCATTGGAAGCTGCA	
DQ191041	CATGGAAAAGAAAAAGGAAGAGAGATATAAGGAAGGTGGTGAAGGGAGAGGCGGTAGCAATGTGTTGGAAGAAATCTTGTGCATTGGAAGCTGCA	
	1401	
AF319776	AGAACATTGCTGACCCATCACACGCCGACATATTC AACCTTAGAGCTGGTCGCGTACGCACCATTAATAGCTTGACCCCTCCCGGTTCTCAAATTGCTC	
AF319777	AGAACATTGCTGACCCATCACACGCCGACATATTC AACCTTAGAGCTGGTCGCGTACGCACCATTAATAGCTTGACCCCTCCCGGTTCTCAAATTGCTC	
AY649099	AGAACATTGCTGACCCATCACACGCCGACATATTC AACCTTAGAGCTGGTCGCGTACGCACCATTAATAGCTTGACCCCTCCCGGTTCTCAAATTGCTC	
DQ191041	AGAACATTGCTGACCCATCACACGCCGACATATTC AACCTTAGAGCTGGTCGCGTACGCACCATTAATAGCTTGACCCCTCCCGGTTCTCAAATTGCTC	
	1501	
AF319776	TCTCAGCGCCCAATGGGTTAAACTCTACAAGGTCACACAAATATATTACCAAAATGTTACTTGCTAGTTGTTGTTATCTTTTCCTTCAATCTTATATAT	
AF319777	TCTCAGCGCCCAATGGGTTAAACTCTACAAG-----	
AY649099	TCTCAGCGCCCAATGGGTTAAACTCTACAAGGTCACACAAATATATTACCAAAATGTTACTTGCTAGTTGTTGTTATCTTTTCCTTCAATCTTATATAT	
DQ191041	TCTCAGCGCCCAATGGGTTAAACTCTACAAG-----	
	1601	
AF319776	ATATA--GTAATTGCATCAGGACCACACTTACAACCACAATAACCGCATTTTCTAACGATCCCGCGACCGCAAGACAATTTAAACCATATTTTATAT	
AF319777	-----	
AY649099	ATATATAGTAATTGCATCAGGACCACACTTACAACCACAATAACCGCATTATTCTAACGATCCCGCGACCGCAAGACAATTTAAACCATATTTTATAT	
DQ191041	-----	



图中阴影核苷酸序列表示外显子,无阴影核苷酸序列表示内含子,下划线表示内含子碱基差异,方框表示外显子碱基差异。
The shadow means that the exons of Gy7,the underling means the difference of bases between introns and the frame means that of exons

图5 Gy7 全基因及 cDNA 测序结果分别与 Genbank(AF319776 和 AF319777) 序列比对
Fig.5 Comparison of the sequences between Gy7,Gy7 cDNA,Genbank AF319776 and AF319777

表1 Genbank(AF319777) 公布的 Gy7 cDNA 和南农 87C-38 大豆 Gy7 cDNA 之间碱基及其对应的氨基酸差异分析
Table 1 Differences of the bases and amino acids in the Gy7 cDNA between Nannong87C-38 soybean and genbank AF319777

图5 显示碱基差异位点 The different base positions in Fig. 5	Genbank (AF319777)		南农 87C-38 Nannong87C-38 soybean		氨基酸改变类型 The type of the change of amino acid
	密码子	氨基酸	密码子	氨基酸	
	Codon	Amino acid	Codon	Amino acid	
937	CC <u>A</u>	Pro	CC <u>G</u>	Pro	同义(synonymous codon)
1153	GA <u>A</u>	Glu	GA <u>G</u>	Glu	同义(synonymous codon)
1232	<u>G</u> AT	Asp	<u>A</u> AT	Asn	错义(missense codon)
1353	G <u>T</u> G	Val	G <u>A</u> G	Glu	错义(missense codon)
1417	CC <u>A</u>	Pro	CC <u>T</u>	Pro	同义(synonymous codon)
1465	AT <u>C</u>	Ile	AT <u>T</u>	Ile	同义(synonymous codon)

表中下划线表示密码子中变化的碱基。The underline means the different base among the codon

3 讨论

由于赖氨酸等必需氨基酸的缺乏导致水稻等谷类作物种子在蛋白质质量方面存在不足。在豆类种子蛋白质中赖氨酸含量较高。因此,在设想利用转基因技术改良禾谷类作物营养品质时,人们自然会选用豆类作物贮藏蛋白基因作为目的基因。刘博林等^[13]曾采用等点聚焦电泳和 SDS 电泳,从四棱豆(*Psophocarpus tetragonolobus* L.)种子中分离到一个

富含赖氨酸的蛋白质。随后,Sun 等^[14]克隆了该基因的 cDNA 序列,并命名为高赖氨酸蛋白基因。该基因编码 158 个氨基酸,其中赖氨酸残基数为 17 个。本研究克隆获得的 11S 大豆球蛋白 Gy7 基因 cDNA,尽管分子量较大,但是它的赖氨酸残基总数为 32 个,比 Sun 等从四棱豆种子中克隆到的基因赖氨酸残基数提高 88.24%。

将南农 87C-38 高蛋白大豆 Gy7 基因 cDNA 与 Genbank 已报道由美国学者克隆的 Gy7 基因 cDNA 序列进行比对,在四个外显子中,第 1、第 2 和第 4 外显子的核苷酸序列完全相同,两个不同品种 Gy7

基因 cDNA 序列所有 6 个核苷酸差异都位于第 3 外显子。由此推测,第 1、第 2 和第 4 外显子编码的氨基酸在 G7 多肽形成特定高级结构中可能具有非常重要的作用,它们在进化过程中受到的选择压力可能比第 3 外显子更强。

近年来,利用基因工程技术将富含赖氨酸的蛋白质基因导入水稻、小麦等作物,从而提高这些作物种子贮藏蛋白质质量方面的研究在国内外都已取得一定进展。在国内,高越峰等^[15]和王逸群等^[16]利用组成型启动子、唐俐等^[17]利用水稻种子特异表达启动子分别引导四棱豆(*Psophocarpus tetragonolobus* L.)高赖氨酸基因在水稻中表达,分别提高了水稻叶片和种子中的赖氨酸含量。孟超敏等^[18]将来自 *E. coli* 的赖氨酸合成关键酶基因(dapA,编码 DHDPS)和来自四棱豆的高赖氨酸基因同时导入小麦,转基因植株后代叶片游离赖氨酸含量有了不同程度的提高。在国外,Zheng 等^[19]、Sindhu 等^[20]和 Katsube 等^[21]先后报道了分别将 β -菜豆蛋白基因(β -Phaseolin)、豌豆球蛋白基因(LegA)和大豆球蛋白基因(AlaB1b,即 Gy3)成功导入水稻改良水稻种子的蛋白质质量。从 Genbank 已登陆的 β -菜豆蛋白基因(β -Phaseolin)(序列号 J01263)编码序列分析, β -Phaseolin 基因总的氨基酸数为 435 个,赖氨酸残基数为 25 个。从 Genbank 已登陆的豌豆球蛋白基因(LegA)(序列号 X02982)编码序列分析,LegA 基因总的氨基酸数为 507 个,赖氨酸残基数为 23 个。从 Genbank 已登陆的大豆球蛋白 Gy3 基因(序列号 X15123)编码序列分析,该基因总的氨基酸数为 481 个,赖氨酸残基数为 19 个。

尽管试验所克隆的 Gy7 基因赖氨酸残基数在 513 个氨基酸残基总数(23 个氨基酸残基为信号肽)中的比例为 6.2,低于 Sun 等人从四棱豆种子中克隆到基因(10.8),但是,据前人研究认为,大豆球蛋白与水稻主要贮藏蛋白中的谷蛋白在氨基酸组成上具有 32% 的同源性^[22],而且两者具有类似的分子结构和类似的翻译后加工、转运、包装程序^[23-26]。作者还曾比较了大豆 Gy7 基因和水稻谷蛋白 Gt1 基因密码子组成,两者密码子的偏爱性也相同(未发表)。另外,从利用 Gy7 基因转化水稻等谷类作物提高赖氨酸含量的绝对值角度考虑,大豆 Gy7 基因可能将是一个更适合于改良水稻等谷类作物营养品质的优质目的基因。

参 考 文 献

[1] 阎其涛,李建粤,米东,等. 重要谷类种子贮藏蛋白的特性及改良研究[J]. 西北植物学报,2004,24(4):754-759.

[2] 孙桂华,王玉凤,崔天明,等. 大麦的营养价值及其开发利用[J]. 辽宁农业科学,1996(3):38-41.

[3] 李小红. 大豆在我国膳食中的应用前景分析[J]. 作物研究,1997,11(1):37-39.

[4] Fukushima D. Recent progress of soybean protein foods; Chemistry, technology, and nutrition [J]. Food Reviews International, 1991,7(3):323-351.

[5] 刘后利. 农作物品质育种[M]. 武汉:湖北科学技术出版社,2001:282.

[6] Beilinson V, Chen Z, Shoemaker C, et al. Genomic organization of glycinin genes in soybean[J]. Theoretical and Applied Genetics, 2002,104(67):1132-1140.

[7] Fischer R L, Goldberg R B. Structure and flanking regions of soybean seed protein genes[J]. Cell, 1982,29:651-660.

[8] Scallan B, Thanh V H, Floener L A, et al. Identification and characterization of DNA clones encoding group - II glycinin subunits [J]. Theoretical and Applied Genetics, 1985,70:510-519.

[9] Beilinson V, Chen Z, Shoemaker C, et al. Genomic organization of glycinin genes in soybean[J]. Theoretical and Applied Genetics, 2002,104(6-7):1132-1140.

[10] 李建粤,米东,张伟,等. 大豆球蛋白研究以及在改良稻米营养品质中的应用. 西北植物学报[J]. 2005,25(8):1706-1712.

[11] 萨姆布鲁克丁,拉塞尔 DW,黄培堂. 分子克隆实验指南(第三版)[M]. 北京:科学出版社,2002,8:27-114.

[12] Nielsen N C, Dickinson C D, Cho T J, et al. Characterization of the glycinin gene family in soybean[J]. Plant Cell, 1989,1(3):313-328.

[13] 刘博林,荆玉祥,匡廷云. 高赖氨酸植物种的筛选及其蛋白质纯化与鉴定[J]. 植物学报,1993,35(1):62-68.

[14] Sun S M, Xing L W, Jing X Y, et al. Lysine rich protein from winged bean[P]. U. S. Patent 6184437, 1998.

[15] 高越峰,荆玉祥,沈世华,等. 高赖氨酸蛋白基因导入水稻及可育转基因植株的获得[J]. 植物学报,2001,43(5):506-511.

[16] 王逸群,郑金贵,谢宝贵. 农杆菌介导将高赖氨酸蛋白基因导入谷秆两用水稻[J]. 应用与环境生物学报,2005,11(3):276-278.

[17] 唐俐,刘巧泉,邓晓湘,等. 无抗性选择标记的转高赖氨酸蛋白(LRP)基因稻株恢复系的获得[J]. 作物学报,2006,32(8):1248-1251.

[18] 孟超敏,陈绪清,梁荣奇,等. 高赖氨酸含量基因在转基因小麦的表达及其赖氨酸含量分析[J]. 科学通报,2004,49(17):1731-1736.

遗传改良,作为辅助选择标记,最终可以有效降低群体亚麻酸含量。而通常标记选择研究中,总是希望得到准确的标记,除了基于基因本身部分序列开发标记外^[9],实际中图谱法得到的标记与目的基因都有约 10 cM 的差异,甚至更大不可能作为准确选择的依据,尤其是单株选择依据。但是可以作为改良群体整体效应的标记,从而有效提高育种效率。

参 考 文 献

[1] 徐杰. 不同品质类型大豆品种脂肪酸形成规律的研究[D]. 东北农业大学硕士学位论文,2005.

[2] Cregan P B, Jarvik T. An integrated genetic linkage map of the soybean genome[J]. Crop Science, 1999, 35(5):1464 – 1490.

[3] 宛煜嵩, 王珍, 方宣钧, 等. 一张含有 227 个 SSR 分子标记的大豆遗传连锁图[J]. 分子植物育种, 2005, 3(1):15 – 20.

[4] 陈庆山, 张忠臣, 刘春燕, 等. 应用 Charleston × 东农 594 重组自交系群体构建 SSR 大豆遗传图谱[J]. 中国农业科学, 2005, 38(7):1312 – 1316.

(上接 846 页)

[19] Zheng Z W, Sumi K, Tanaka K, et al. The bean seed storage protein β -Phaseolin is synthesized, processed, and accumulated in the vacuolar Type – II protein bodies of transgenic rice endosperm[J]. Plant Physiology, 1995, 109:777 – 786.

[20] Sindhu A S, Zheng Z W, Murai H. The pea seed storage protein legumin was synthesized, processed, and accumulated stably in transgenic rice endosperm[J]. Plant Science, 1997, 130(2):189 – 196.

[21] Katsube T, Kurisaka N, Ogawa M, et al. Accumulation of Soybean Glycinin and Its Assembly with the Glutelins in Rice[J]. Plant Physiology, 1999, 120:1063 – 1073.

[22] Higuchi W, Fukazawa C. A rice glutelin and a soybean glycinin have evolved from a common ancestral gene[J]. Gene, 1987, 55

[5] Pirjo Tanhuanpää, Alan Schulman. Mapping of genes affecting linolenic acid content in *Brassica rapa ssp. oleifera*. [J]. Molecular Breeding, 2002, 10:51 – 62.

[6] Schneider, KA, Brothers M E, Kelly J D. Marker assisted selection to improve drought resistance in common bean[J]. Crop Science, 1997, 37:51 – 60.

[7] Rahman S M, Kinoshita T, Anai, et al. Combining ability in loci for high oleic and low linolenic acids in Soybean[J]. Crop Science, 2001, 41:26 – 29.

[8] 杨柳, 张彬彬, 韩英鹏, 等. 大豆亚麻酸 QTL 的分析[J]. 大豆科学, 2006, 25(3):270 – 274.

[9] 薛庆中, 张能义, 熊兆飞. 应用分子标记辅助选择培育抗白叶枯病水稻恢复系[J]. 浙江农业大学学报, 1998, 24(6):631 – 638.

[10] Song Q J, Shoemaker R C. A new integrated genetic linkage map of soybean[J]. Theoretical and Applied Genetics, 2004, 109:122 – 128.

[11] Shah S, Xin Z, Browse J. Overexpression of the FAD3 desaturase Gene in a mutant of arabidopsis[J]. Plant Physiology, 1997, 114:1533 – 1539.

(2 – 3):245 – 253.

[23] Yamagata H, Sugimoto T, Tanaka K, et al. Biosynthesis of storage proteins in developing rice seeds[J]. Plant Physiology, 1982, 70:1094 – 1100.

[24] Zhao W M, Gatehouse J A, Boulter D. The purification and partial amino acid sequence of a polypeptide from the glutelin fraction of rice grains; homology to pea legumin. FEBS Letters, 1983, 162:96 – 102.

[25] Wen T N, Luthe D S. Biochemical characterization of rice glutelin[J]. Plant Physiology, 1985, 78:172 – 177.

[26] Okita T W, Hwang Y S, Hnilo J H, et al. Structure and expression of the rice glutelin multigene family[J]. Journal of biological chemistry, 1989, 264(21):12573 – 12581.