



大豆 2S 清蛋白基因生物信息学分析

王迪^{1,2}, 邱红梅², 王曙明², 高淑芹², 马晓萍², 王洋², 郑大浩¹, 王跃强²

(1. 延边大学, 吉林 延吉 133000; 2. 吉林省农业科学院 大豆研究所/大豆国家工程研究中心, 吉林 长春 130033)

摘要: 2S 清蛋白含有动物胰蛋白酶抑制剂及过敏原等多种具有生理活性的多肽, 可引起特殊人群尤其是儿童的过敏反应, 其编码基因为 LTP 基因家族。为系统地分析 2S 清蛋白编码基因, 利用生物信息学分析方法获得 2S 清蛋白基因的全序列、定位以及转录本等信息, 鉴定了 30 个大豆 LTP 家族基因。基因定位结果表明: 23.33% 基因在染色体上串联排布, 反映了该家族基因的串联重复进化。通过基因功能注释得到 22 个非特异性脂质转运蛋白, 1 个类伸展素蛋白, 4 个脯氨酸蛋白, 2 个种子储存 2S 清蛋白和 1 个雄蕊特殊蛋白。根据系统发育分析将这些蛋白分成 3 个类群。类群 I 有类伸展素蛋白、脯氨酸蛋白、雄蕊特殊蛋白和种子储存 2S 清蛋白, 类群 II 有 16 个非特异性脂质转运蛋白, 类群 III 有 6 个非特异性脂肪转移蛋白。种子储存 2S 清蛋白基因在种子发育中后期特异表达, 表达量显著高于其它同源基因。本研究为进一步研究其功能提供指导, 为该类基因应用于大豆品质改良提供理论支撑。

关键词: 大豆; LTP; 进化分析; 差异表达

Comparative Analysis of Soybean 2S Albumin Gene

WANG Di^{1,2}, QIU Hong-mei², WANG Shu-ming², GAO Shu-qin², MA Xiao-ping², WANG Yang², ZHENG Da-hao¹, WANG Yue-qiang²

(1. Yanbian University, Yanji 133000, China; 2. Soybean Research Institute/National Engineering Research Center for Soybean, Jilin Academy of Agricultural Sciences, Changchun 130033, China)

Abstract: 2S albumin derived from the LTP gene family contains various physiologically active peptides such as animal trypsin inhibitors and allergens, which can cause allergic reactions in special populations, especially in children. In order to systematically analyze the encoding gene of 2S albumin, this study used bioinformatics analysis to obtain the full sequence, localization and transcript information of the 2S albumin gene, identified 30 soybean LTP family genes. The results of gene mapping indicated that 23.33% of the genes were arranged in tandem on the chromosome, reflecting the tandem repeat evolution of the family genes. And there are 22 non-specific lipid transporters, 1 extensin-like protein, 4 proline proteins, 2 seeds, 2S albumin and 1 stamen-specific proteins were obtained by gene function annotation. These proteins were divided into three groups based on phylogenetic analysis. Group I has extensin-like proteins, proline proteins, stamen-specific proteins and seeds to store 2S albumin, group II has 16 non-specific lipid transporters, and group III has 6 non-specific fat transfer proteins. The 2S albumin gene was specifically expressed in the late stage of seed development, and its expression level was significantly higher than other homologous genes, which provided guidance for further study of its function and theoretical support for the application of this gene in soybean quality improvement.

Keywords: Soybean; LTP; Evolutionary analysis; Differential expression

大豆是高蛋白作物, 平均蛋白含量约为 40%, 高于花生、油菜、芝麻等作物^[1]。应用超速离心方法, 按照沉降模式, 在 0.5 离子强度的介质溶液中, 大豆蛋白可分为 2S、7S、11S 和 15S 4 个主要组分, 分别称为 2S 清蛋白、7S 碱性蛋白、11S 球蛋白和 15S 组分蛋白^[2]。各大豆蛋白组分的空间结构不同, 11S 球蛋白、7S 碱性球蛋白分别以六聚体、三聚体和四聚体形式存在; 2S 清蛋白的空间结构松散,

由 α -螺旋和无规卷曲构成; 15S 主要为 7S 多聚体^[3]。7S 碱性球蛋白和 11S 球蛋白占大豆蛋白的 70%^[4-5]; 2S 清蛋白约占总蛋白的 10%, 蛋氨酸含量较丰富; 15S 含量约为 5%。同时 2S 清蛋白含有动物胰蛋白酶抑制剂活性成分, 库尼兹 (Kunitz) 和鲍曼-贝氏 (Bowman-Birk) 分别占大豆脱脂粉重的 1.4% 和 0.6%, 具有对大豆蛋白营养有重要意义的生物活性^[6-7], 同时含有 α -淀粉酶抑制剂和过敏原

收稿日期: 2018-08-09

基金项目: 七大农作物育种专项 (2016YFD0100201); 国家自然科学基金 (31671712); 吉林省科技发展规划 (20160414049GH)。

第一作者简介: 王迪 (1993-), 男, 硕士, 主要从事大豆基因挖掘研究。E-mail: 549869052@qq.com。

通讯作者: 郑大浩 (1963-), 男, 博士, 教授, 主要从事作物遗传与育种研究。E-mail: ljzdhdh@163.com;

王跃强 (1968-), 男, 博士, 研究员, 主要从事大豆资源与遗传育种研究。E-mail: 825164942@qq.com。

等多种具有生理活性的多肽,影响蛋白的营养价值,可引起特殊人群尤其是儿童的过敏反应^[8-10]。大豆 2S 清蛋白的研究对大豆育种、种植、加工等方面具有重要意义,有助于通过现代生物技术及工程原理提高大豆产量、质量。

近年来有关大豆籽粒贮藏蛋白的两种主要的组分 7S 球蛋白和 11S 球蛋白各亚基表达调控的分子机制有大量研究^[11],而 2S 蛋白的探索比较少,2S 清蛋白在大豆中所占比例小,分子量也很少^[12]。2S 清蛋白基因来源于 LTP 基因家族,其生物信息学分析能够为该类基因的定位、分类、组织表达分析、表达调控研究与利用提供研究基础。

1 材料与方法

1.1 材料

在蛋白质家族数据库 (Pfam, <http://pfam.xfam.org/>) 以利用 2S 清蛋白的基因编号“Glyma. 13G288100”为关键词进行检索,获得该基因家族的 ID 编号“PF00234”。在大豆基因组数据库 (Phytozome, <https://phytozome.jgi.doe.gov/>) 中检索 PF00234,获得 2S 清蛋白的 30 个基因的序列。

1.2 方法

1.2.1 基因注释及结构分析 利用下载的氨基酸序列,通过 Phytozome (Phytozome, <https://phytozome.jgi.doe.gov/>) 搜索基因定位,获得 Wm82. a2. v1 和 Wm82. a1 版本基因位点号,下载基因结构。得到基因组长度,mRNA 长度,编码长度,蛋白长度及蛋白种类功能注释。根据数据构建 5′-3′的基因结构和已注释的基因对目标基因进行注释。利用 Phytozome 获得基因结构,并利用下载的氨基酸序列,采用 Clastal X (<http://www.genome.jp/tools/clustalw/>) 进行多重序列比对,结果导入 MEGA 5.0 软件,以邻接法 (Neighbor-Joining, NJ) 构建系统进化树,程序值 (Bootstrap) 设为 1 000,其它参数为默认值。

1.2.2 氨基酸组成分析 利用 DNAMAN 软件的蛋白菜单下的属性及序列组成工具,对 30 条氨基酸序列进行分析,获得了每个基因中 20 种氨基酸的含量、等电点、信号肽长度、分子量。将 8 种人体必需氨基酸 (赖氨酸、色氨酸、苯丙氨酸、蛋氨酸、苏氨酸、异亮氨酸、亮氨酸、缬氨酸) 的含量列表进行比较分析。

1.2.3 基因转录表达分析 在 Phytozome 数据库中找到 30 个基因座 (Wm82. a2. v1 版),然后利用大豆遗传学和基因组学数据库 (Soybase, [\[base.org/sbt/\]\(https://soybase.org/sbt/\)\) 的转录组分析工具 \(<https://soybase.org/soyseq/index.php#>\) 对基因进行表达分析,获得基因在叶、花、莢、根、种子不同发育时期的表达数据,其中种子分别为开花后 25,28,35 和 42 d 取得的种子,记为种子 4、种子 5、种子 6 和种子 7。](https://soy-</p></div><div data-bbox=)

2 结果与分析

2.1 大豆 2S 清蛋白基因成员

由表 1 可知,大豆基因组数据库中检索到 30 个基因座属于 2S 家族。Gm01 染色体上存在 5 个基因,其中 Glyma. 01G130000、Glyma. 01G130100 和 Glyma. 01G130200 基因串联,Glyma. 01G132500 和 Glyma. 01G132600 两个基因串联,Glyma. 01G132600 基因编码蛋白长度最长,为 122 bp,Glyma. 01G130100 基因编码蛋白长度最短,为 111 bp。Gm03 染色体上存在 6 个基因,Glyma. 03G035600 在染色体上方,编码蛋白长度最长,为 130 bp;Glyma. 03G040000 和 Glyma. 03G040100、Glyma. 03G040300 和 Glyma. 03G040400 分别在 2 号染色体上串联,且两组串联基因长度相似;Glyma. 03G040000 基因编码蛋白长度最短,为 117 bp。Gm03 号染色体上存在两个基因 Glyma. 05G057500 和 Glyma. 05G057700,其中 Glyma. 05G057500 编码蛋白长度最长,为 136 bp,Glyma. 05G057700 基因编码蛋白长度最短,为 126 bp。Gm04 号染色体上只存在 1 个基因 Glyma. 08G161700,编码蛋白长度为 139 bp,Gm05 号染色体上存在 5 个基因,其中只有 Glyma. 09G055000 和 Glyma. 09G055100 串联;Glyma. 09G275700 基因编码蛋白最长,为 128 bp;Glyma. 09G055000 和 Glyma. 09G055100 编码蛋白长度相同,为 120 bp;Glyma. 09G278000 编码蛋白长度最短,为 116 bp。Gm6、7、8、9 和 10 号染色体上均只有 1 个基因,分别为 Glyma. 12G213600,Glyma. 13G288100,Glyma. 14G032800,Glyma. 15G265900 和 Glyma. 16G197100,编码的蛋白长度为分别 155,158,119,115 和 93 bp。Gm11 号染色体上有 3 个基因,Glyma. 17G140200、Glyma. 17G140300 和 Glyma. 17G140400 串联且蛋白长度相同,为 131 bp。Gm12 号染色体上有 3 个基因,其中 Glyma. 18G214500 和 Glyma. 18G214600 基因串联;Glyma. 18G214500 编码蛋白长度最长,为 125 bp;Glyma. 18G210900 基因编码的蛋白长度最短,为 117 bp。基因定位结果表明,2S 清蛋白的 30 个基因位于 12 条染色体上,共有 7 个串联,23. 33% 基因在染色体上串联排布,反映了该家族基因的串联重复进化。

表 1 大豆 2S 基因家族成员
Table 1 Summary of 2S family members in soybean

基因位点 Gene site		基因定位 Gene position	基因组长度 Genomic length/bp	蛋白长度 Protein length/bp
Wm82. a2. v1	Wm82. a1			
Glyma. 01G130000	Glyma01g32140	Chr01:44435404.. 44440876 (5. 47 kb)	3909	119
Glyma. 01G130100	Glyma01g32160	Chr01:44470591.. 44472807 (2. 22 kb)	1583	111
Glyma. 01G130200	Glyma01g32190	Chr01:44486020.. 44489059 (3. 04 kb)	2172	117
Glyma. 01G132500	Glyma01g32750	Chr01:45184550.. 45186731 (2. 18 kb)	1558	118
Glyma. 01G132600	Glyma01g32760	Chr01:45213072.. 45215403 (2. 33 kb)	1666	122
Glyma. 03G035600	Glyma03g04370	Chr03:4294866.. 4296708 (1. 84 kb)	1317	130
Glyma. 03G040000	Glyma03g04910	Chr03:5008635.. 5011258 (2. 63 kb)	1874	117
Glyma. 03G040100	Glyma03g04920	Chr03:5024735.. 5027318 (2. 58 kb)	1846	122
Glyma. 03G040300	Glyma03g04950	Chr03:5037967.. 5040758 (2. 79 kb)	1994	118
Glyma. 03G040400	Glyma03g04960	Chr03:5050094.. 5052676 (2. 58 kb)	1845	122
Glyma. 03G257200	Glyma03g41800	Chr03:45186230.. 45187144 (916 bp)	653	118
Glyma. 05G057500	Glyma05g04440	Chr05:5331131.. 5332832 (1. 7 kb)	1216	136
Glyma. 05G057700	Glyma05g04460	Chr05:5345406.. 5345938 (534 bp)	381	126
Glyma. 08G161700	Glyma08g17190	Chr08:12569634.. 12571927 (2. 29 kb)	1638	139
Glyma. 09G055000	Glyma09g06100	Chr09:4891035.. 4892390 (1. 36 kb)	968	120
Glyma. 09G055100	Glyma09g06130	Chr09:4905206.. 4906649 (1. 45 kb)	1032	120
Glyma. 09G275500	Glyma09g41200	Chr09:49105606.. 49107220 (1. 62 kb)	1153	125
Glyma. 09G275700	Glyma09g41220	Chr09:49113820.. 49115260 (1. 44 kb)	1029	128
Glyma. 09G278000	Glyma09g41460	Chr09:49336167.. 49338084 (1. 92 kb)	1370	116
Glyma. 12G213600	Glyma12g34160	Chr12:37245803.. 37246851 (1. 05 kb)	749	155
Glyma. 13G288100	Glyma13g36400	Chr13:38876766.. 38877998 (1. 23 kb)	881	158
Glyma. 14G032800	Glyma14g03660	Chr14:2382959.. 2383947 (990 bp)	707	119
Glyma. 15G265900	Glyma15g42000	Chr15:50106991.. 50109805 (2. 82 kb)	2011	115
Glyma. 16G197100	Glyma16g32170	Chr16:35851181.. 35852418 (1. 24 kb)	884	93
Glyma. 17G140200	Glyma17g14910	Chr17:11394144.. 11395557 (1. 42 kb)	1010	131
Glyma. 17G140300	Glyma17g14921	Chr17:11398220.. 11399323 (1. 1 kb)	788	131
Glyma. 17G140400	Glyma17g14930	Chr17:11401941.. 11403264 (1. 32 kb)	946	131
Glyma. 18G210900	Glyma18g44300	Chr18:49735182.. 49737498 (2. 32 kb)	1655	117
Glyma. 18G214500	Glyma18g44670	Chr18:50124652.. 50127460 (2. 81 kb)	2007	125
Glyma. 18G214600	Glyma18g44680	Chr18:50129946.. 50131519 (1. 57 kb)	1124	124

2.2 LTP 基因家族结构分析

根据对 2S 清蛋白家族的 30 个基因的结构分析,基因长度为 380 bp 到4 000 bp 不等,Gm05 的 12、13 与 17 号染色体上共 7 个基因均只有 1 个外显子无内含子。Gm01 号染色体上 Glyma. 01G130400 基因无 5'与 3'UTR 结构。系统地分析鉴定这 30 个大豆 LTP 家族基因。通过 Soybase

网站的基因功能注释查找后,该 LTP 家族含有 5 种蛋白,22 个非特异性脂质转运蛋白,1 个类伸展素蛋白,4 个脯氨酸蛋白,2 个种子储存 2S 清蛋白和 1 个雄蕊特异性蛋白。分别占总体的 73. 33%、3. 33%、13. 33%、6. 66% 和 3. 33%,非特异性脂质转运蛋白最多,脯氨酸蛋白次之(表 2)。

表 2 大豆 2S 家族基因注释及结构分析

Table 2 Gene annotation and structural analysis of soybean 2S family

基因编号 Gene number	mRNA 长度 mRNA length/bp	编码区长度 Coding area length/bp	蛋白种类 Protein type	基因结构(5'-3') Gene structure(5'-3')
Glyma.01G130000	614	360	非特异性脂质转运蛋白	
Glyma.01G130100	778	369	非特异性脂质转运蛋白	
Glyma.01G130200	806	393	非特异性脂质转运蛋白	
Glyma.01G130300	720	375	非特异性脂质转运蛋白	
Glyma.01G130400	381	381	非特异性脂质转运蛋白	
Glyma.03G035600	679	387	非特异性脂质转运蛋白	
Glyma.03G040000	625	363	非特异性脂质转运蛋白	
Glyma.03G040100	766	351	非特异性脂质转运蛋白	
Glyma.03G040300	946	396	非特异性脂质转运蛋白	
Glyma.03G040400	1010	396	非特异性脂质转运蛋白	
Glyma.03G257200	922	378	非特异性脂质转运蛋白	
Glyma.05G057500	788	396	类伸展素蛋白	
Glyma.05G057700	757	357	富含脯氨酸的蛋白质	
Glyma.08G161700	578	363	非特异性脂质转运蛋白	
Glyma.09G055000	881	477	非特异性脂质转运蛋白	
Glyma.09G055100	1484	348	非特异性脂质转运蛋白	
Glyma.09G275500	734	378	非特异性脂质转运蛋白	
Glyma.09G275700	749	468	非特异性脂质转运蛋白	
Glyma.09G278000	608	354	非特异性脂质转运蛋白	
Glyma.12G213600	566	282	种子储存 2S 清蛋白	
Glyma.13G288100	336	336	种子储存 2S 清蛋白	
Glyma.14G032800	646	354	非特异性脂质转运蛋白	
Glyma.15G265900	791	354	非特异性脂质转运蛋白	
Glyma.16G197100	809	369	雄蕊特异性蛋白	
Glyma.17G140200	839	369	富含脯氨酸的蛋白质	
Glyma.17G140300	709	360	富含脯氨酸的蛋白质	
Glyma.17G140400	958	420	富含脯氨酸的蛋白质	
Glyma.18G210900	1216	411	非特异性脂质转运蛋白	
Glyma.18G214500	953	357	非特异性脂质转运蛋白	
Glyma.18G214600	535	357	非特异性脂质转运蛋白	

➡:基因方向;█:外显子;—:内含子;□:5'与3'UTR。

➡:Gene direction;█:Representative exon;—:Representing introns;□:5' and 3' UTR.

2.3 LTP 家族蛋白功能及进化分析

大豆 2S 清蛋白的 30 个基因可按照蛋白质功能分为蛋白酶抑制剂/种子储藏/LTP 家族、双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白、非特异性脂肪转移蛋白 8、非特异性脂肪转移蛋

白 15、蛋白酶抑制剂/种子储藏/LTP 家族相关蛋白、双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白。其中蛋白酶抑制剂/种子储藏/LTP 家族的基因有 18 个,双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白的基因有 5 个,非

特异性脂肪转移蛋白 8 的基因有 2 个,非特异性脂肪转移蛋白 15 的基因有 2 个,蛋白酶抑制剂/种子储藏/LTP 家族相关蛋白的基因有 1 个,双功能的抑

制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白的基因有 2 个(表 3)。

表 3 功能分类

Table 3 Functional classification

基因编号 Gene number	功能分类 Functional classification
Glyma. 01G130000	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 01G130100	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma0. 1G130200	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 01G130300	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 01G130400	双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白
Glyma. 03G035600	双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白
Glyma. 03G040000	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 03G040100	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 03G040300	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 03G040400	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 03G257200	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 05G057500	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 05G057700	双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白
Glyma. 08G161700	非特异性脂肪转移蛋白 8
Glyma. 09G055000	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 09G055100	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 09G275500	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 09G275700	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 09G278000	非特异性脂肪转移蛋白 15
Glyma. 12G213600	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 13G288100	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 14G032800	蛋白酶抑制剂/种子储藏/LTP 家族
Glyma. 15G265900	非特异性脂肪转移蛋白 8
Glyma. 16G197100	蛋白酶抑制剂/种子储藏/LTP 家族相关蛋白
Glyma. 17G140200	双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白
Glyma. 17G140300	双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白
Glyma. 17G140400	双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白
Glyma. 18G210900	非特异性脂肪转移蛋白 15
Glyma. 18G214500	双功能的抑制剂/脂肪转移蛋白/种子储存 2S 清蛋白相关蛋白
Glyma. 18G214600	蛋白酶抑制剂/种子储藏/LTP 家族

根据系统发育分析将这些蛋白分成 3 个类群。类群 I 有类伸展素蛋白、脯氨酸蛋白、雄蕊特殊蛋白和种子储存 2S 清蛋白,类群 II 有 16 个非特异性脂

质转运蛋白,类群 III 有 6 个非特异性脂肪转移蛋白(图 1)。

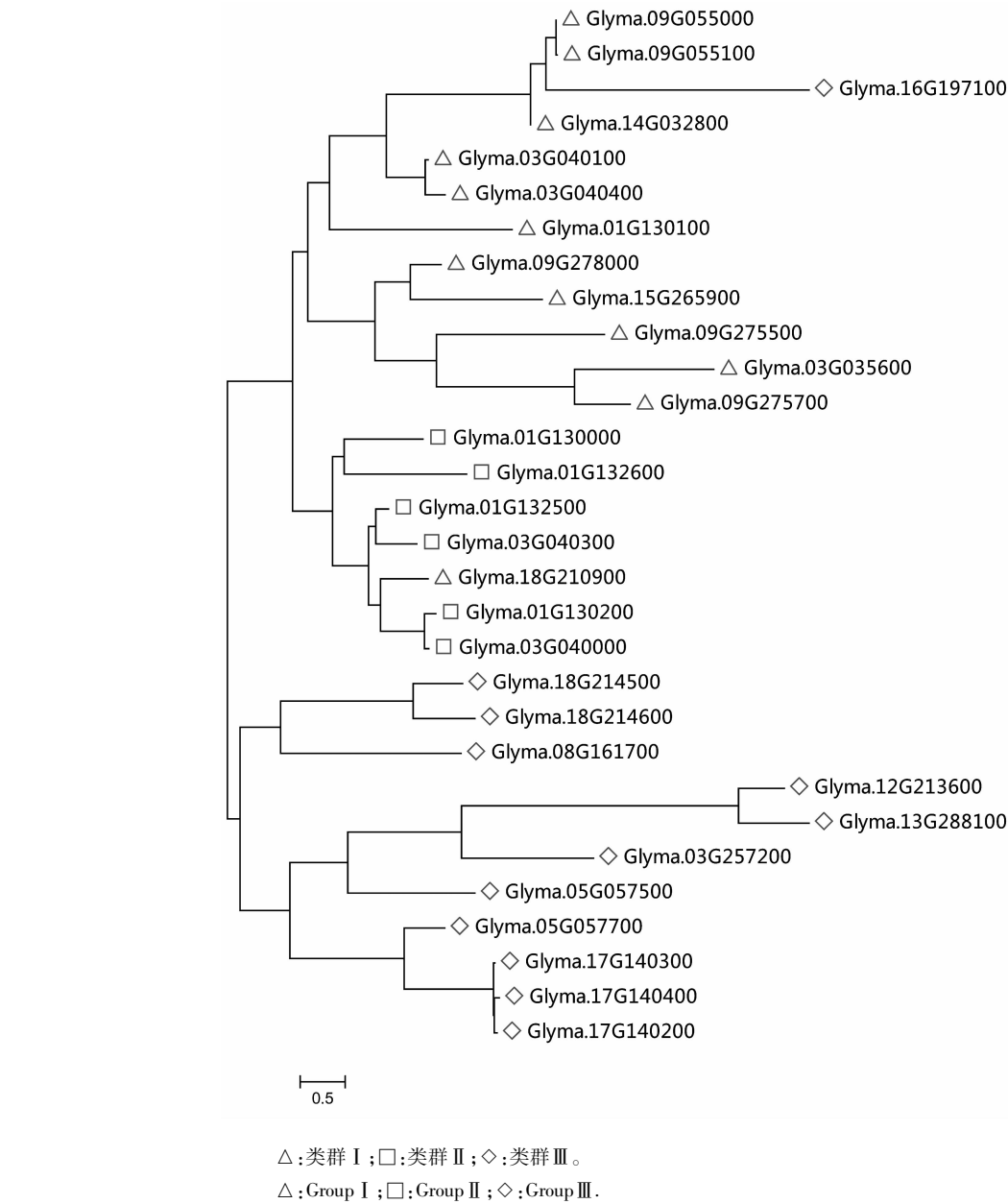


图 1 LTP 家族蛋白进化树

Fig. 1 LTP family protein evolution tree

2.4 LTP 家族蛋白性质及必需氨基酸组成

分别对系统进化树划分的 3 类 LTP 家族蛋白的性质和必需氨基酸组成进行比较分析,其中 76.67% 的基因编码序列具有信号肽(表 4)。I 类中 2 个种子储存 2S 清蛋白均含信号肽,信号肽长度 16 ~ 20 aa,其中多数蛋白的等电点偏酸性。该类蛋白中苯丙氨酸、缬氨酸和苏氨酸均含量极低,蛋氨酸含量为 18.22% ~ 19.33%,在植物种子中特异性表达。I 类中的 1 个类伸展素蛋白含有信号肽,信号肽长度为 25 aa,等电点偏中性。该类蛋白中蛋氨酸和色氨酸含量为 0,亮氨酸含量为 45.79%,此类蛋白在荚皮和根中表达。此外 I 类中还存在 4 个脯氨酸蛋白,这 4 个脯氨酸蛋白信号肽长度均为 21 aa,其中 3 个偏碱性,这 3 个脯氨酸蛋白不含有蛋氨酸

和色氨酸,剩余的 1 个脯氨酸偏酸性,在各个器官中都不表达,这一蛋白不含有亮氨酸和缬氨酸。II 类中有 16 个非特异性脂质转运蛋白,其中 6 个没有信号肽,余下 10 个基因信号肽的长度为 18 ~ 48 aa。这 10 个有信号肽的蛋白中 1 个偏酸性,1 个偏中性,8 个为碱性。不含有信号肽的蛋白等电点均偏碱性,这 6 个不含有信号肽的蛋白在植物中均不表达。II 类的 16 种蛋白中只有 2 个蛋白含有少量色氨酸,其它 14 个非特异性脂质转运蛋白不含色氨酸。III 类中的 6 个蛋白均含有信号肽,其中有 2 个偏酸性,1 个偏中性,其它蛋白均为碱性,信号肽长度为 21 ~ 47 aa,异亮氨酸含量较少,亮氨酸含量偏高。

表 4 8 种氨基酸含量
Table 4 8 kinds of amino acid content

基因编号 Gene number	信号肽 长度 Signal peptide length/bp	成熟蛋 白长度 Mature protein length/bp	分子量 Molecular weight	等电点 Isoelectric point	氨基酸 Amino acid							
					Phe	Ile	Lys	Leu	MET	Thr	Val	Trp
Glyma. 01G130000	0	119	12581. 3	10. 06	1. 12	6. 24	1. 99	6. 24	3. 04	6. 48	8. 76	0
Glyma. 01G130100	0	111	12298. 2	11. 37	3. 47	5. 51	4. 09	7. 35	1. 04	5. 84	8. 20	1. 43
Glyma. 01G130200	0	117	12217. 0	9. 89	2. 31	5. 50	6. 13	7. 33	4. 17	5. 00	9. 82	0
Glyma. 01G130300	25	93	9312. 7	8. 99	1. 51	8. 37	9. 33	5. 98	0	9. 77	4. 27	0
Glyma. 01G130400	25	97	10596. 1	8. 19	2. 68	5. 32	4. 74	5. 32	2. 42	5. 80	8. 55	0
Glyma. 03G035600	24	106	11355. 9	8. 20	2. 49	6. 93	3. 31	5. 94	2. 25	6. 29	7. 96	0
Glyma. 03G040000	48	69	7400. 1	10. 21	1. 92	6. 08	3. 39	3. 04	1. 73	8. 28	9. 51	0
Glyma. 03G040100	28	94	9481. 9	8. 99	2. 96	4. 70	9. 17	7. 05	2. 67	6. 40	3. 15	0
Glyma. 03G040300	25	93	10082. 0	11. 81	1. 41	5. 59	4. 98	7. 82	0	6. 09	6. 98	0
Glyma. 03G040400	0	122	12365. 0	8. 73	2. 27	3. 61	8. 04	6. 31	8. 20	6. 55	8. 05	0
Glyma. 03G257200	21	97	10187. 8	7. 63	2. 77	7. 70	8. 59	6. 60	2. 50	7. 00	4. 91	0
Glyma. 05G057500	25	111	11555. 4	7. 85	1. 22	2. 91	4. 32	17. 44	0	4. 40	7. 79	0
Glyma. 05G057700	21	105	10917. 0	5. 57	5. 28	12. 57	15. 38	0	7. 23	10. 07	0	2. 83
Glyma. 08G161700	47	92	9112. 6	8. 20	0	8. 54	9. 52	7. 32	0	4. 43	5. 45	0
Glyma. 09G055000	21	99	10514. 1	5. 15	4. 04	6. 41	13. 09	5. 34	1. 21	3. 88	2. 86	0
Glyma. 09G0551000	21	99	10542. 1	5. 15	4. 03	6. 39	13. 06	5. 33	1. 21	3. 87	3. 81	0
Glyma. 09G275500	0	125	12963. 0	8. 84	1. 09	6. 90	6. 73	11. 22	3. 93	7. 05	9. 25	0
Glyma. 09G275700	25	103	11549. 4	7. 38	3. 70	6. 86	3. 28	6. 86	5. 57	6. 23	6. 12	0
Glyma. 09G278000	18	98	10475. 0	6. 20	2. 70	4. 29	4. 78	6. 44	3. 66	3. 90	8. 62	0
Glyma. 12G213600	20	135	15660. 2	5. 70	0	7. 98	10. 51	7. 26	6. 60	0. 66	0	1. 13
Glyma. 13G288100	16	142	16647. 8	4. 85	0. 86	5. 47	10. 66	6. 15	6. 22	3. 10	0. 61	1. 06
Glyma. 14G032800	21	98	10331. 0	8. 59	1. 37	5. 43	16. 94	5. 43	0	7. 89	6. 79	0
Glyma. 15G265900	23	92	9112. 6	8. 20	0	8. 54	9. 52	7. 32	0	4. 43	5. 45	0
Glyma. 16G197100	0	93	9671. 3	5. 55	1. 46	4. 63	1. 29	13. 89	3. 95	7. 36	7. 24	0
Glyma. 17G140200	21	110	11497. 6	9. 02	2. 45	3. 90	13. 03	13. 64	0	5. 31	9. 57	0
Glyma. 17G140300	21	110	11530. 6	8. 83	2. 45	3. 89	11. 91	14. 58	0	4. 41	9. 55	0
Glyma. 17G140400	21	110	11532. 6	8. 83	2. 45	2. 92	11. 91	13. 61	0	4. 41	11. 28	0
Glyma. 18G210900	20	97	10423. 1	8. 40	2. 72	4. 32	6. 01	6. 48	6. 14	7. 84	7. 71	0
Glyma. 18G214500	23	102	10806. 4	8. 98	2. 62	6. 23	8. 10	5. 19	0	8. 49	9. 28	0
Glyma. 18G214600	0	124	13020. 0	8. 84	2. 17	7. 75	5. 76	10. 33	2. 94	6. 25	9. 22	1. 34

2.5 LTP 家族基因的转录表达

LTP 家族基因在不同组织部位的转录分析结果显示:Gm01 染色体上第 1 个串联位点的 3 个基因均无表达,并且不含有信号肽,第 2 个串联位点的 2 个基因均表达且其中 1 个高表达;Gm03 染色体上 6 个基因均不表达;Gm05 染色体上有 2 个基因均属于 I 类群,一个在荚皮和根中高表达,另一个在各个位置都不表达;Gm08 染色体上的 1 个基因在各部位

都不表达;Gm09 染色体上的 1 处串联的 2 个基因不表达,其余 3 个中有一个在花中表达,一个不表达,一个在花叶荚皮中高表达;第 6 个染色体上有 1 个基因,是 I 类群中种子储存 2S 清蛋白基因,在种子中有高表达,而且有信号肽;Gm12 染色体上的 1 个基因在种子中有高表达,而且有信号肽;Gm13 和 Gm14 染色体上均有 1 个基因含有信号肽,不表达;Gm15 染色体上有 1 个基因,不含信号肽,不表达;

Gm16 染色体上有 3 个基因串联在一起,这 3 个基因信号肽长度、成熟蛋白长度、分子量等电点以及各个氨基酸含量基本相同,且这 3 个基因在根中高表达;Gm17 染色体上有 3 个基因,其中 2 个串联在一起,这 3 个基因都在花中高表达。在大豆种子发育过程中高表达的基因有 Glyma. 01G132500、Glyma. 12G213600 和 Glyma. 13G288100。2S 清蛋白基

因在种子发育中后期特异表达,表达量显著高于其它基因。3 个脯氨酸蛋白基因在根中特异表达。类伸展素蛋白基因只在荚皮和根中表达。II 类群的 4 个基因只在花中表达,2 个基因在叶、花、荚、种子中都高表达;此外,还有 1 个基因在叶、花和荚中高表达(表 5)。

表 5 不同组织部位基因的表达

Table 5 Gene expression of different tissue parts								
基因 Gene	叶 Leaf	花 Flower	荚 Pod	种子 1 Seed1	种子 2 Seed2	种子 3 Seed3	种子 4 Seed4	根 Root
Glyma. 01G130000	73	37	68	0	0	0	0	0
Glyma. 01G130100	0	0	0	0	0	0	0	0
Glyma. 01G130200	0	0	0	2	1	4	0	0
Glyma. 01G132500	5550	4351	5899	1970	2499	783	227	4
Glyma. 01G132600	277	381	77	23	9	21	7	0
Glyma. 03G035600	266	123	21	32	27	37	20	0
Glyma. 03G040000	0	0	0	0	0	0	0	0
Glyma. 03G040100	1	113	11	4	6	4	3	0
Glyma. 03G040300	73	73	23	0	0	0	0	0
Glyma. 03G040400	1	13	2	144	161	199	66	0
Glyma. 03G257200	0	0	0	2	1	0	1	0
Glyma. 05G057500	1	27	296	0	0	1	0	2203
Glyma. 05G057700	0	1	0	0	0	0	0	1
Glyma. 08G161700	1	7	1	0	0	0	0	516
Glyma. 09G055000	0	0	1	23	7	9	2	0
Glyma. 09G055100	0	0	1	15	5	17	3	0
Glyma. 09G275500	0	2395	0	0	0	0	0	0
Glyma. 09G275700	0	53	0	0	0	0	0	0
Glyma. 09G278000	727	3412	2758	0	0	0	0	4
Glyma. 12G213600	0	0	0	1820	5631	28850	32384	0
Glyma. 13G288100	0	0	0	3451	11092	47269	65571	0
Glyma. 14G032800	0	0	2	175	90	109	13	0
Glyma. 15G265900	0	1	0	0	0	0	0	290
Glyma. 16G197100	0	0	0	0	0	0	0	0
Glyma. 17G140200	4	49	156	1	0	0	1	2963
Glyma. 17G140300	2	21	2	0	0	0	1	2963
Glyma. 17G140400	0	3	2	0	0	0	0	2961
Glyma. 18G210900	806	1487	473	146	77	69	15	3
Glyma. 18G214500	0	3261	1	6	7	4	0	0
Glyma. 18G214600	0	1079	2	38	80	25	0	0

3 结论与讨论

与传统的基于实验室的生物学研究方法相比,生物信息学方法可用较低的成本和较快的时间获得较可靠的数据结果^[13]。大豆种子贮藏蛋白的遗传改良研究已经在种质筛选、组分及亚基的结构与功能、亚基的遗传规律、遗传工程和诱变育种等方面进行了广泛而深入的研究,并取得了一系列的成

果^[14]。通过对大豆 2S 清蛋白初步分析了解到大豆 2S 清蛋白分子量小,蛋白占比也很少,仅有 10%。但是它分子结构较为稳定,可以抵抗蛋白水解,还能抵抗化学诱导变性,并且具有较高的抗蛋白酶消化能力。

本文通过对 30 个 2S 家族基因进行 LTP 家族的进化分析,结果显示 LTP 是一种多功能的蛋白质,根据基因功能注释将 30 个基因分为 5 个基因功能

家族,鉴定分析了 30 个基因的表达位置,筛选特异性表达基因,为性状改良深入研究打下了基础。根据进化树分析将蛋白分为 3 类,确定了每个基因的 8 种氨基酸含量,等电点等。虽然 30 个基因在 12 条染色体上都有分布,但是只有部分基因表达且特异性较高。这些分析结果为接下来的基因调控途径研究提供了新的思路,为基因功能研究分析提供了理论依据。

参考文献

[1] 裴剑慧,王强,周素梅. 我国花生蛋白资源的开发与利用[J]. 粮油加工与食品机械,2005(12):47-49. (Pei J H, Wang Q, Zhou S M. Development and utilization of peanut protein resources in China[J]. Cereals and Oils Processing, 2005(12):47-49.)

[2] Peng I C , Q uass D W , Dayton W R, et al .The physicochemical and functional properties of soybean 11S globulin a review[J]. Cereal Chemistry, 1984 , 61(6) : 480-490.

[3] 邹谋勇. 大豆贮藏蛋白序列分析与 3-D 结构同源模建[J]. 食品与发酵科技, 2016, 52(4) : 25-32. (Zou M Y. Sequence analysis and 3D structure homology modeling of soybean storage proteins[J]. Food and Fermentation Sciences & Technology, 2016, 52(4) : 25-32.)

[4] Tarone A G, Fasolin L H, Perrechil F DA, et al. Influence of drying conditions on the gelling properties of the 7S and 11S soy protein fractions [J]. Food and Bioproducts Processing, 2013, 91(2) : 111-120.

[5] Zhao X Y, Chen F S, Xue W D, et al. FTIR spectra studies on the secondary strcutures of 7S and 11S globulins from soybean proteins using AOT reverse micellar extraction[J]. Food Hydrocoloids, 2008, 22(4) :568-575.

[6] Koshiyama I, Kikuchi M, Harada K, et al. 2S globulins of soybean seeds. 1. isolation and characterization of protein components [J]. Journal of Agricultural and Food Chemistry, 1981, 29(2) : 336-340.

[7] 周瑞宝, 周兵. 大豆 7S 和 11S 球蛋白的结构和功能性质[J].

中国粮油学报,1998(6):41-44. (Zhou R B, Zhou B. The structure and functional properties of soybean 7S and 11S globulin proteins[J]. Journal of the Chinese Cereals and Oils Association, 1998(6) :41-44.)

[8] 赵京,陈育智,文昭明. 婴幼儿和儿童食物过敏的发病机理及临床表现[J]. 中华儿科杂志, 2002(3) : 54-56. (Zhao J, Chen Y Z, Wen S M. Pathogenesis and clinical manifestations of food allergies in infants and children[J]. Chinese Journal of Pediatrics, 2002(3) : 54-56.)

[9] 胡声迪,郭金枝,马曦. 大豆 β -伴大豆球蛋白研究进展[J]. 中国畜牧杂志, 2010, 46(5) : 65-68. (Hu S D, Guo J Z, Ma X. Advance in soybean β -conglycinin[J]. Chinese Journal of Animal Science, 2010, 46(5) : 65-68.)

[10] Han Y S, Lin J, Bardina L, et al. What characteristics confer proteins the ability to induce allergic responses? IgE epitope mapping and comparison of the structure of soybean 2S albumins and Ar ah 2[J]. Molecules, 2016, 21(5) : 622.

[11] 刘珊珊,武小霞,姜振峰,等. 大豆 7S 球蛋白 β -伴大豆球蛋白的研究现状[J]. 大豆科学, 2007,26(3) : 417-422. (Liu S S, Wu X X, Jiang Z F, et al. Advance in soybean 7S globulin β -conglycinin[J]. Soybean Science, 2007, 26(3) :417-422.)

[12] 张崇本,吴显荣. 我国大豆种子中球蛋白 2S 组分的分离纯化及部分性质的研究[J]. 生物化学杂志,1991(2) : 230-236. (Zhang C B, Wu X R. The seperation, purification and some properties of 2S fraction of soybean seed globulins [J]. Chinese Journal of Biochemistry and Molecular Biology, 1991(2) : 230-236.)

[13] 吕笑言,金英. 利用生物信息技术构建分析甜菜 M14 品系盐胁迫下参考转录组数据库[J]. 黑龙江大学自然科学学报, 2017,34(2) : 208-212. (Lyu X Y, Jin Y. Bioinformatics analysis of the reference transcriptome in sugar beet M14 line under salt stress[J]. Journal of Natural Science of Heilongjiang University, 2017, 34(2) : 208-212.)

[14] 刘春,王显生,麻浩. 大豆种子贮藏蛋白遗传改良研究进展[J]. 大豆科学,2008,27(5) :866-873. (Liu C, Wang X S, Ma H. Genetic improvement on soybean seed storage proteins[J]. Soybean Science,2008,27(5) :866-873.)