

大豆 CYS 家族基因的鉴定、启动子分析及 EST 表达研究

谢翎^{1,2}, 金正俊², 汪章勋³, 黄勃³

(1. 安庆师范学院 生命科学院, 安徽 安庆 246011; 2. 韩国国家农业科学院 农业生物部, 京畿道 水原 441-707, 韩国; 3. 安徽农业大学 微生物防治省重点实验室, 安徽 合肥 230036)

摘要: 利用大豆基因组数据库, 通过生物信息学分析, 鉴定并获得大豆所有 CYS 家族基因的序列、结构以及定位信息。全基因组分析检测表明, 大豆染色体 DNA 中共含有 19 个 CYS 家族基因成员。基因定位分析表明, 这些基因分别分布于大豆的 12 条染色体。系统进化分析将大豆的全部 CYS 基因分成了 3 个亚家族。启动子分析揭示, 19 个大豆 CYS 基因的上游调控序列都含有逆境反应顺式元件。EST 表达谱分析表明, 大豆 CYS 家族基因大多在子叶、花、叶片以及种皮等组织中优势表达。

关键词: CYS 家族基因; 差异表达; 进化分析

中图分类号: S565.1

文献标识码: A

DOI: 10.11861/j.issn.1000-9841.2014.05.0654

Genome-wide Identification, EST Expression, and Upstream Sequence Analysis of CYS Family Genes in Soybean

XIE Ling^{1,2}, KIM Jeong-jun², WANG Zhang-xun³, HUANG Bo³

(1. School of Life Sciences, Anqing Teachers College, Anqing 246011, China; 2. Agricultural Microbiology Division, National Academy of Agricultural Science, RDA, Sumon 441-707, Korea; 3. Anhui Provincial Key Laboratory for Microbial Pest Control, Anhui Agricultural University, Hefei 230036, China)

Abstract: Based on bioinformatic method, the CYS family genes of soybean were identified and their sequences, structures and chromosome location were determined. CYS proteins were classified according to their structure and phylogeny relationship. The results of genome analysis showed that a total of 19 CYS genes, located on 12 chromosomes, respectively, could be systematically identified from soybean genome and classified into three classes. All of them had stressed related cis-acting elements in their upstream sequences. The EST data of CYS family genes from the UniGene database was used to analyze the expression pattern of those genes in different tissues. EST analysis revealed that there was predominant expression of most soybean CYS family genes in cotyledon, flower, leaf and seed coat. These results will facilitate functional analysis and utilization of CYS genes in soybean.

Key words: Soybean; CYS family genes; Differential expression; Phylogenetic analysis

蛋白酶抑制剂根据作用底物的不同, 可以分为丝氨酸蛋白酶抑制剂、天冬氨酸蛋白酶抑制剂、金属蛋白酶抑制剂以及半胱氨酸蛋白酶抑制剂 4 种类型^[1]。其中, 关于植物半胱氨酸蛋白酶抑制剂 (Cystain) 的研究起步较晚。由于害虫及病原菌在有害农作物时, 其作用酶主要为半胱氨酸蛋白酶, 理论上植物内源的 Cystain 在一定程度上可以降低病虫害的影响。因此, CYS 基因家族的相关研究, 对于揭示植物自身病虫害防御作用机制具有重要的意义。另有报道表明, 相同植物物种中编码形成不同 Cystain 的功能基因 (CYS 基因), 可能是由同一相关基因进化而来^[2], 进而形成了 CYS 基因家族。

目前, 在拟南芥 (*Arabidopsis thaliana*)、水稻 (*Oryza sativa*) 及大麦 (*Hordeum vulgare*) 等诸多植物

中, 都已经发现有 Cystain 的存在^[3-5]。现有的研究表明, 植物体内的半胱氨酸蛋白酶抑制剂基因 CYS 与植物防御病虫害确实有重要关联。如拟南芥 CYS 基因家族成员, 在受到侵染时其编码产物会表现出明显的对病原微生物和害虫的抗性^[3]; 水稻中的 CYS 基因的表达可抑制稻瘟病菌 (*Magnaporthe oryzae*) 的生长, 并对鞘翅目害虫表现出抗性^[6]。

大豆 (*Glycine max*) 作为一种重要的经济作物, 其 CYS 家族基因的相关研究还非常匮乏。本研究利用生物信息学手段, 鉴定了大豆基因组中全部 CYS 基因家族成员, 并从基因水平上分析 CYS 基因在大豆中的染色体定位、基因结构、调控元件、以及进化与表达模式, 以期今后为进一步在大豆中研究和利用 CYS 基因奠定基础。

收稿日期: 2014-01-21

基金项目: 国家自然科学基金 (31201568); 安徽省高校省级自然科学基金项目 (KJ2011B081, KJ2013B123); 安庆师范学院科研启动项目 (044-K05000000281)。

第一作者简介: 谢翎 (1980-), 女, 博士, 主要从事生物化学与分子生物学研究。E-mail: xxve@sohu.com。

通讯作者: 黄勃 (1969-), 男, 博士, 教授, 主要从事分子生物学研究。E-mail: bohuan@ahau.edu.cn。

1 材料与方法

1.1 材料

研究所使用的大豆相关编码区序列数据及基因组数据均来源于 NCBI 数据库(<http://www.ncbi.nlm.nih.gov/>)和 Phytozome 网站大豆信息数据库(<http://www.phytozome.org/soybean.php>)。拟南芥与水稻的 CYS 基因数据和蛋白数据分别来源于 TAIR(<http://arabidopsis.org/>)^[7]和 TIGR(<http://rice.plantbiology.msu.edu/>)^[8]数据平台。大豆 EST 表达数据来源于 UniGene 数据平台。

1.2 大豆 CYS 基因家族成员的识别和结构域分析

使用模式植物拟南芥和水稻基因组中已分类鉴定的所有 CYS 家族基因成员,利用 Pfam 数据库工具^[9]获得多种相关的典型蛋白结构域。以拟南芥和水稻 CYS 蛋白典型结构域序列 tBLASTN($P = 0.001$)搜索大豆的全基因组数据库,获得大豆相关的同源序列,同时手工剪切无完整读码框序列。使用 SMART 数据库工具^[10],进一步分析初筛出的大豆相关同源序列结构域,并剔除无典型 CYS 蛋白结构域的序列,得到大豆中所有 CYS 编码基因。利用 Yang 等^[11]的研究方法判别同源基因,即两基因联配序列的相似性大于 70%,且联配部分的长度应覆盖两个基因中较长基因的 80% 的序列。使用 GSDS 工具(<http://gsds.cbi.pku.edu.cn/>)^[12]绘制所有大豆 CYS 基因的基因结构图。

1.3 大豆 CYS 基因家族各成员在染色体上的物理定位

利用 perl 程序解析大豆基因组信息,确定大豆 CYS 基因成员在各条染色体上的具体位置。按照各基因成员和不同染色体的长度比例利用 MapDraw V2.1 软件^[13]绘制大豆 CYS 基因染色体定位图。

1.4 大豆 CYS 基因家族编码蛋白的多序列联配和系统发育分析

对大豆 CYS 家族基因成员的各编码蛋白序列

使用 BioEdit V7.0.1 软件进行联配分析。使用 MEGA V5.1 软件^[14]对联配结果采用邻接法生成 CYS 基因的系统进化树,校验参数 bootstrap 值设置为重复 1 000 次。

1.5 大豆 CYS 家族基因上游启动子分析

由大豆基因组数据库获取大豆 CYS 基因家族成员编码区上游序列(2 000 bp),作为启动子调控序列进行分析。利用植物顺式作用元件分析软件 PLACE 和 PlantCARE 分析并统计大豆 CYS 基因家族成员的启动子作用元件。

1.6 大豆 CYS 基因家族成员的组织表达分析

将已获取的所有大豆 CYS 基因家族成员所对应的编码区序列分别与 NCBI 数据平台大豆 EST 数据库进行 tblastn 比对,取 $E \leq 10^{-10}$ 、匹配率 $\geq 95\%$,且序列长度 ≥ 100 bp 的结果作为对应大豆的 CYS 基因 EST 序列。对 UniGene 中检索到的大豆各 CYS 基因 EST 表达数据进行器官组织来源分类,并分析结果。

2 结果与分析

2.1 大豆 CYS 家族基因鉴定与分类

通过全基因组数据检测,由大豆基因组 DNA 中分离子得到了 19 个 CYS 基因家族(表 1)。根据 CYS 基因家族所在染色体的相关信息,对这 19 个基因分别进行命名,从 *GmCYS1* 到 *GmCYS19*。结构域分析,发现这 19 个 CYS 基因家族成员编码蛋白均含有 CYS 典型特征结构域。大豆 CYS 基因家族成员编码蛋白生化属性分析表明,最短的大豆 CYS 蛋白(编码基因 *GmCYS9* 和 *GmCYS19*)仅含有 97 个氨基酸,而最长蛋白(编码基因 *GmCYS10* 和 *GmCYS15*)包含了 245 个氨基酸残基。大豆 CYS 基因家族成员编码蛋白的等电点范围从 4.98(编码基因 *GmCYS13*)到 9.88(编码基因 *GmCYS1*、*GmCYS5*、*GmCYS6* 和 *GmCYS18*)。

表 1 大豆 CYS 家族基因基本信息

Table 1 Basic information of CYS family genes in soybean

基因 Gene	基因序列号 Gene ID number	DNA 长度 DNA length/bp	编码区长度 CDS length/bp	Cystatin 结构域位置 Cystatin domain location(Pfam)	氨基酸 Amino acid	等电点 Isoelectric point	分子量 Molecular mass/Da
<i>GmCYS1</i>	Glyma04g10360	720	345	26 ~ 107	114	9.88	12645.5
<i>GmCYS2</i>	Glyma05g28250	962	393	33 ~ 98	130	9.23	14182.4
<i>GmCYS3</i>	Glyma07g39590	1270	345	27 ~ 87	114	7.77	12745.7
<i>GmCYS4</i>	Glyma08g11210	773	312	4 ~ 56	103	8.76	11725.5
<i>GmCYS5</i>	Glyma09g01360	345	345	26 ~ 110	114	9.88	12703.6
<i>GmCYS6</i>	Glyma09g16960	670	345	26 ~ 94	114	9.88	12629.5
<i>GmCYS7</i>	Glyma11g06850	345	345	26 ~ 107	114	9.83	12721.6

续表 1

基因	基因序列号	DNA 长度	编码区长度	Cystatin 结构域位置	氨基酸	等电点	分子量
Gene	Gene ID number	DNA length/bp	CDS length/bp	Cystatin domain location(Pfam)	Amino acid	Isoelectric point	Molecular mass/Da
<i>GmCYS8</i>	Glyma11g36780	975	381	27 ~ 100	126	9.52	13991.1
<i>GmCYS9</i>	Glyma13g04250	3449	294	14 ~ 92	97	5.83	10681.1
<i>GmCYS10</i>	Glyma13g25870	4084	738	49 ~ 135	245	6.56	27668.0
<i>GmCYS11</i>	Glyma14g04250	1652	312	5 ~ 91	103	5.8	11409.1
<i>GmCYS12</i>	Glyma14g04260	1589	603	11 ~ 97	200	5.26	22531.6
				106 ~ 189			
<i>GmCYS13</i>	Glyma14g04291	1930	555	11 ~ 86	184	4.98	20771.5
				101 ~ 173			
<i>GmCYS14</i>	Glyma15g12211	1021	339	25 ~ 110	112	9.07	12415.3
<i>GmCYS15</i>	Glyma15g36180	4650	738	50 ~ 135	245	7.27	27619.0
				156 ~ 220			
<i>GmCYS16</i>	Glyma18g00690	781	429	30 ~ 97	142	9.56	15695.1
<i>GmCYS17</i>	Glyma18g12240	2426	363	5 ~ 91	120	6.29	13667.7
<i>GmCYS18</i>	Glyma19g39400	1050	345	26 ~ 107	114	9.88	12645.5
<i>GmCYS19</i>	Glyma20g08800	3303	294	14 ~ 92	97	5.83	10741.1

2.2 大豆 CYS 家族基因的基因结构分析

基因结构分析表明(图 1),大豆的 CYS 基因家族各成员基因长度具有明显的差异。其中最短的 *GmCYS5* 和 *GmCYS7* 只有 345 bp;而 *GmCYS15* 序列最长,达到了 4 650 bp。此外,大豆 CYS 家族基因的内含子数目也较少,介于 0 ~ 3 个,长度为 89 ~ 1 540 bp;其外显子数目为 1 ~ 4 个,长度介于 62 ~ 429 bp。其中,19 个大豆 CYS 家族基因中有 11 个

成员的结构基因不含有内含子序列,占成员总数的 58%。在含有内含子序列的 8 个大豆 CYS 基因中,各有 3 个基因分别含有 1 和 2 个内含子,均占成员总数的 16%;另外,还有 2 个基因含有 3 个内含子,约占成员总数的 10%。大豆 CYS 基因家族各成员基因的外显子-内含子结构异化较为严重,可能会影响其基因功能的分化。

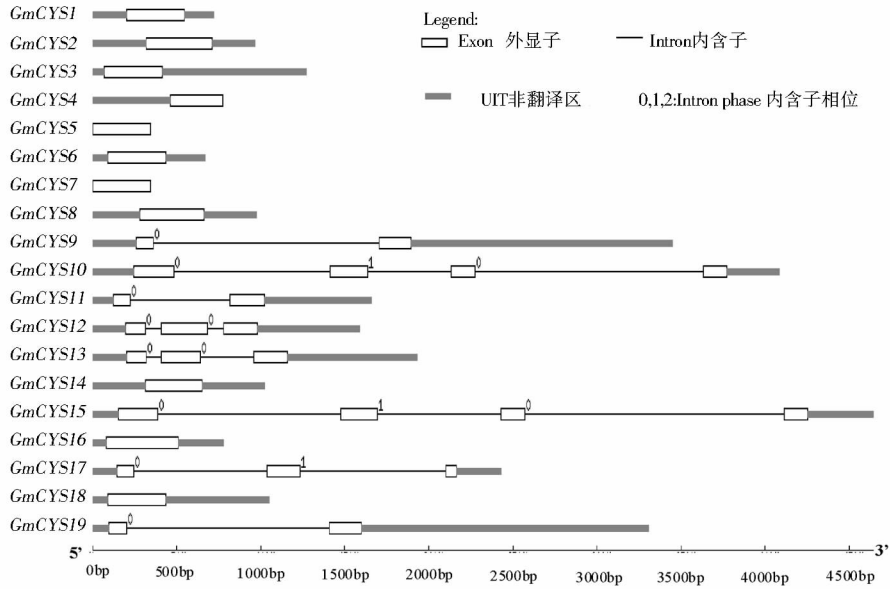


图 1 大豆 CYS 基因家族的基因结构

Fig. 1 The gene structure of CYS family genes in soybean

2.3 大豆 CYS 家族基因染色体定位

为确定大豆 CYS 家族基因各成员在不同染色体上的具体分布,利用 Phytozome 数据平台中相关基因位置信息,绘制了大豆各 CYS 基因的染色体定位图(图 2)。19 个 CYS 家族基因,分布于大豆 12 条染色体上。其中,第 4、5、7、8、19 和 20 号染色体

上含有 1 个 CYS 基因;第 9、11、13、15 和 18 号染色体上含有 2 个 CYS 基因;第 14 号染色体上含有 3 个 CYS 基因;而其余各染色体则不含 CYS 基因。此外,染色体定位还表明绝大多数大豆 CYS 基因成员都靠近所在染色体的两端。

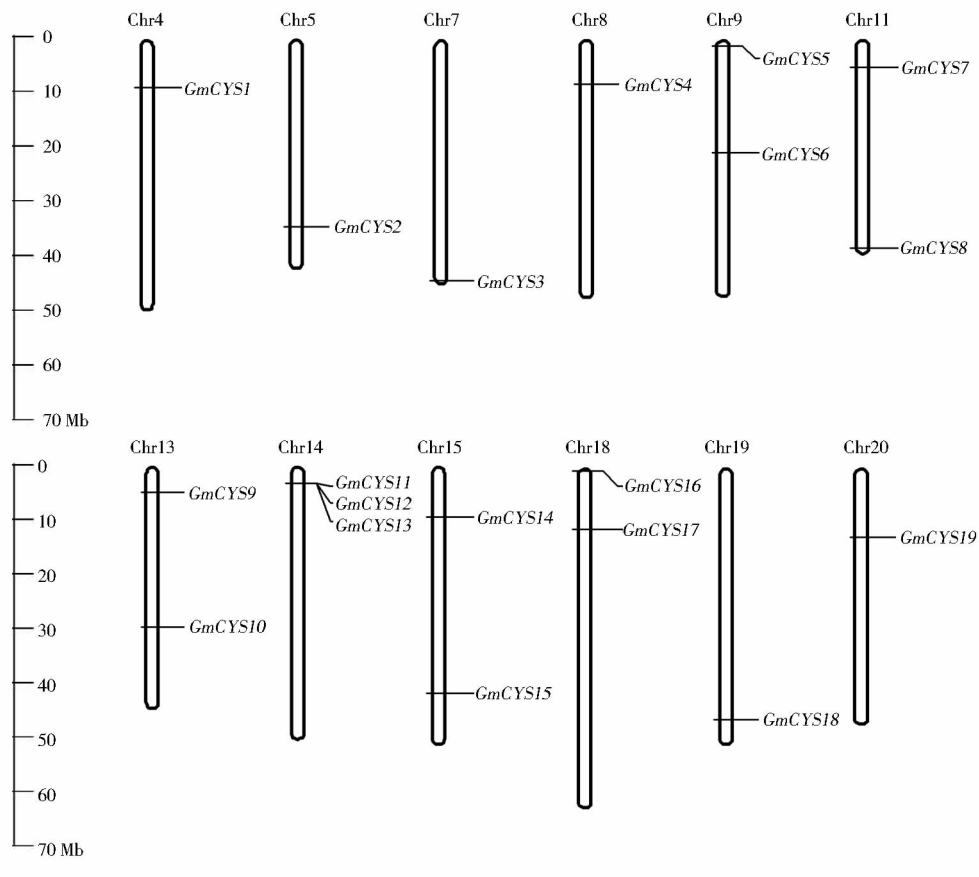


图 2 大豆 *CYS* 基因的染色体分布

Fig.2 Chromosome distribution of *CYS* genes in soybean genome

2.4 大豆 *CYS* 家族基因的系统进化及保守蛋白结构域分析

为明确大豆 *CYS* 基因家族成员间的系统进化

关系,使用大豆各 *CYS* 基因编码蛋白序列构建系统进化树。结果表明(图 3),19 个大豆 *CYS* 基因共分为 3 类。参照水稻与大麦相关研究结果^[5],将 3 个

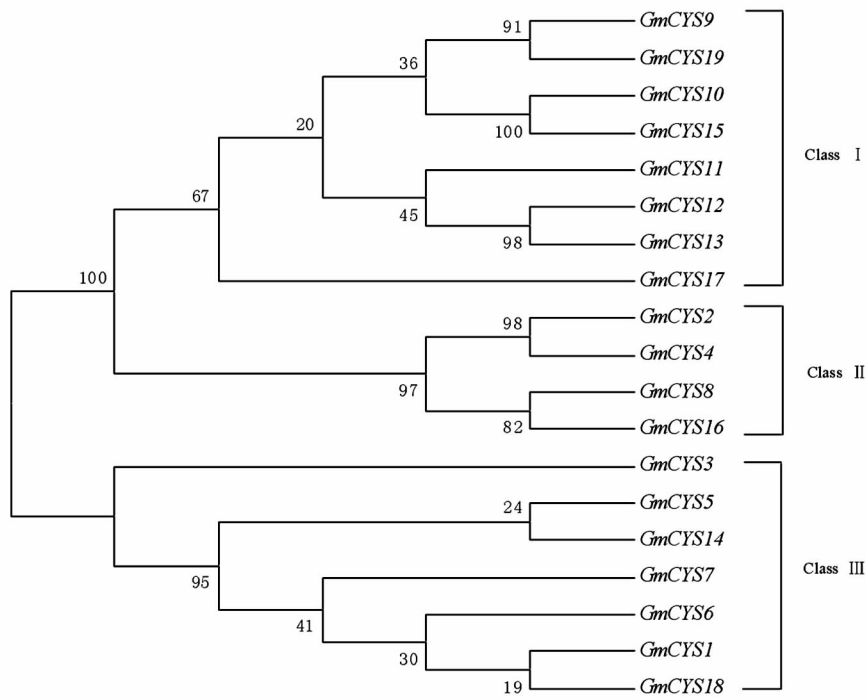


图 3 大豆 *CYS* 基因家族编码蛋白邻接系统树

Fig.3 Neighbor-joining phylogenetic tree of the *CYS* in soybean

大豆 CYS 基因亚家族命名为 Class I、Class II 和 Class III, 它们分别含有 8, 4 和 7 个 CYS 基因。另外, 19 个 CYS 基因形成了 7 个旁系基因对, 其中 4 对基因 bootstrap 值 (bootstrap values) 大于 90, 包括: *GmCYS9/GmCYS19*、*GmCYS10/GmCYS15*、*GmCYS12/GmCYS13* 和 *GmCYS2/GmCYS4*。

同时, Class I 中的 8 个 CYS 家族成员都含有内含子 (1~3 个); 而 Class II 和 Class III 中的 CYS 家族基因则均不含内含子 (图 1)。蛋白序列系统进化与基因结构分析表明, 大豆 CYS 家族基因中聚类关系比较近的成员其基因结构也非常相似, 如 *GmCYS10* 和 *GmCYS15*, *GmCYS9* 和 *GmCYS19*, *GmCYS12* 和 *GmCYS13* 均分别聚类在一起 (图 1 和图 3)。这可能表明内含子-外显子结构一定程度上在大豆 CYS 家族成员中起到了系统发生关系指数的作用。

具有高保守的半胱氨酸蛋白酶抑制剂功能域 (Pfam: Cystatin domain) 是植物 CYS 蛋白最重要的特征^[15]。PFAM 和 SMART 结构域分析显示, 19 个大豆 CYS 家族基因编码蛋白均含有 CYS 典型特征结构功能域。其中, *GmCYS12*、*GmCYS13*、*GmCYS15* 的编码蛋白序列中, 含有 2 个 Cystatin domain; 其余

大豆 CYS 蛋白则只含有 1 个 Cystatin domain (表 1)。大豆各 CYS 蛋白的结构域长度也表现出比较大的差异, *GmCYS4* 编码蛋白的 Cystatin domain 最短, 仅 53 个氨基酸, 而最长的大豆 CYS 蛋白 Cystatin domain 则达到了 87 个氨基酸 (*GmCYS10*、*GmCYS11*、*GmCYS12-1* 及 *GmCYS17*)。另外, 大多数大豆 CYS 成员的特征结构域在其对应的编码蛋白序列中占据了较高的比例, 如蛋白 *GmCYS11* 的结构域在其编码蛋白序列中所占的比例接近 85%。

将大豆 CYS 基因 3 个亚家族成员编码蛋白结构域中不能对齐的序列剪切后分别进行联配比对分析 (图 4)。其中 Class II、Class III 的 CYS 蛋白功能域较为保守, 而 Class I 中 8 个 CYS 蛋白的 11 个 Cystatin domain 保守性则相对较低。另外, 在 3 个 CYS 亚家族各成员的编码蛋白中均发现了高度保守的氨基酸序列 QVvxG。该序列是 CYS 蛋白的重要活性位点。该活性位点的第 4 个氨基酸在 Class II 四个成员中均统一编码为丝氨酸; 在 Class III 成员中主要编码为丙氨酸 (*GmCYS5*、*GmCYS14* 编码丝氨酸和缬氨酸); 而在 Class I 成员中的多样性更高, 可编码为丝氨酸、丙氨酸、谷氨酸以及天冬氨酸。

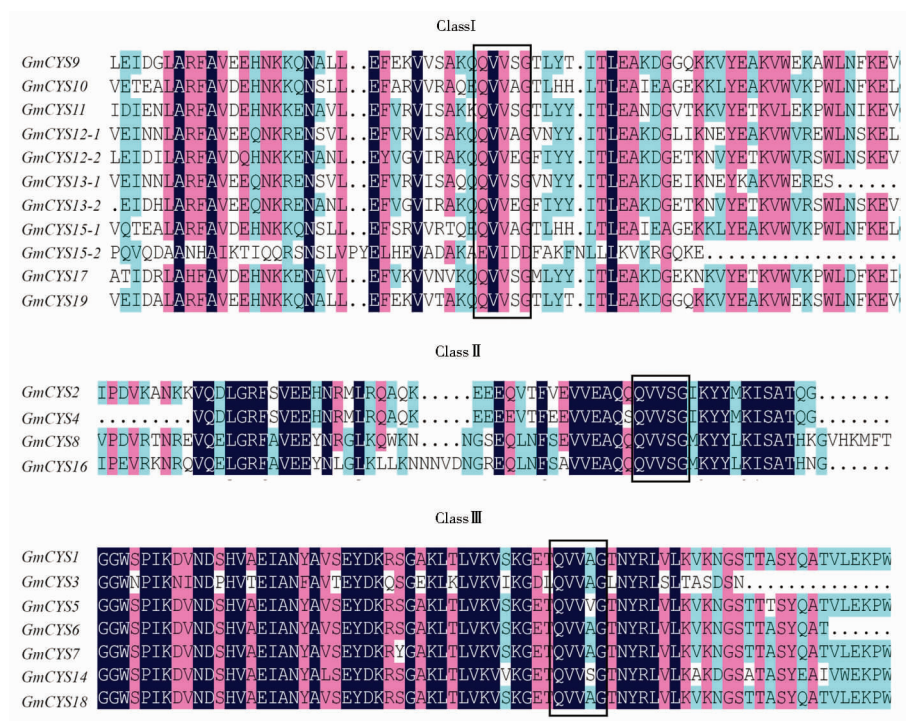


图 4 大豆 CYS 蛋白不同亚家族主要结构域蛋白序列比对

Fig. 4 Alignment of conserved motifs of different class of the CYS in soybean

2.5 大豆 CYS 家族基因的启动子序列分析

大豆各 CYS 基因启动子序列分析显示, 19 个 *GmCYS* 基因的启动子调控区域均含有大量与干旱或 ABA 诱导表达有关的转录因子作用元件, 如: MYB 和 MYC。此外, 各大豆 CYS 基因上游启动子

序列中, 还包含诸多与信号分子响应和环境因子应答有关的元件。统计表明, 有 89.5% 的大豆 CYS 基因成员含有脱落酸响应元件 (ABRE); 78.9% 的 CYS 基因含防卫和胁迫响应元件 (TC-rich repeats); 73.7% 的基因具有热响应元件 (HSE); 63.2% 的大

豆 CYS 基因具有茉莉酸甲酯反应元件(TGACG-motif);另有 78.9% 的 CYS 基因含有节律控制应答元件(circadian)。该结果表明,各 CYS 基因的表达与大豆逆境生存具有重要的内在联系。

表 2 大豆 CYS 家族基因上游序列中相关元件数量分布

Table 2 Distribution of relative elements in upstream sequences of soybean CYS genes

基因 Gene	脱落酸 响应元件 ABRE	热激 响应元件 HSE	茉莉酸甲酯 反应元件 TGACG-motif	防卫和胁迫 响应元件 TC-rich repeats	节律控制 应答元件 circadian	MYB 结合元件 MYB binding element	MYC 结合元件 MYC binding element
GmCYS1	2	2	2	2	3	13	9
GmCYS2	1	0	2	3	2	12	10
GmCYS3	3	1	0	4	4	8	6
GmCYS4	1	2	0	1	1	12	6
GmCYS5	2	1	1	2	1	6	9
GmCYS6	2	0	2	2	1	11	5
GmCYS7	3	1	2	2	1	15	6
GmCYS8	2	2	0	1	0	14	5
GmCYS9	3	0	3	0	2	10	9
GmCYS10	0	1	2	2	1	10	6
GmCYS11	5	1	2	0	1	12	7
GmCYS12	1	3	1	3	1	10	7
GmCYS13	3	2	0	3	2	5	9
GmCYS14	2	4	0	1	0	13	6
GmCYS15	2	1	1	3	2	8	6
GmCYS16	4	0	0	3	0	16	3
GmCYS17	0	0	0	0	0	3	7
GmCYS18	3	1	2	0	2	13	7
GmCYS19	2	4	1	5	1	18	8

2.6 大豆 CYS 基因家族成员基因表达分析

利用 UniGene 数据库对大豆 CYS 基因家族各成员进行 EST 表达分析(表 3)。在 19 个大豆 CYS 基因中,有 5 个基因未找到 EST 证据,分别为 GmCYS1、GmCYS5、GmCYS7、GmCYS13 以及 GmCYS18。在 14 个含有 EST 表达数据的大豆 CYS 基因中,有组织特异性表达的分别为 GmCYS2、GmCYS4、GmCYS8 和 GmCYS16。其中,GmCYS2 和 GmCYS4 在种皮特异表达;

而 GmCYS8 和 GmCYS16 则在子叶特异表达。值得注意的是,4 个特异性表达的大豆 CYS 基因均来自第二亚家族成员,且特异表达部位相同的基因,其编码蛋白在进化分析时也聚在了一起。其他 10 个大豆 CYS 基因在多个器官中均有表达。如:GmCYS10、GmCYS11、GmCYS12、GmCYS14 以及 GmCYS15 表达的组织范围较广,且表达谱较为接近,在子叶、花、叶片、分生组织以及种皮和胚中都有表达。

表 3 大豆不同组织中 CYS 基因的 EST 表达分析

Table 3 EST expression analysis of soybean CYS genes in different tissues

组织 Tissue	子叶 Cotyledon	花 Flower	叶片 Leaf	分生组织 Meristem	果荚 Pod	根 Root	种皮 Seed coat	胚 Embryo	茎 Stem	叶芽 Vegetative bud
GmCYS2	-	-	-	-	-	-	181	-	-	-
GmCYS3	32	45	62	-	228	21	271	-	-	250
GmCYS4	-	-	-	-	-	-	22	-	-	-
GmCYS6	95	31	21	-	-	62	71	-	-	-
GmCYS8	487	-	-	-	-	-	-	-	-	-
GmCYS9	64	101	21	-	114	-	-	-	-	-
GmCYS10	116	367	196	116	-	371	351	482	-	119
GmCYS11	32	831	544	498	456	-	181	91	94	322
GmCYS12	129	741	417	316	374	-	198	119	81	186
GmCYS14	194	415	65	62	-	696	90	148	94	150
GmCYS15	129	118	196	166	-	65	362	370	47	-
GmCYS16	64	-	-	-	-	-	-	-	-	-
GmCYS17	41	51	41	-	-	61	-	-	-	-
GmCYS19	55	72	18	-	97	-	-	-	-	-

“-”表示无匹配 EST 数据;表达值单位:每百万转录本中基因的 EST 数。
“-”:No matching EST data was found;Expression unit;EST transcripts per million(TPM).

3 讨 论

研究认为,基因家族是由一个祖先基因通过重复进化而得到的一系列功能和序列非常相似的基因。由于大豆的全基因组序列测序已经完成,这就为在基因组水平上分析大豆编码基因的功能提供了重要的手段。作为广泛存在于植物物种中,能够有效抑制某些害虫和病原微生物体内的半胱氨酸蛋白酶活性的 *CYS* 基因,人们普遍认为它是植物自身病虫害防御体系中的重要一环。

根据目前现有报道,在几个重要植物物种里,拟南芥和小麦中均发现了 7 个 *CYS* 基因,水稻中发现了 12 个 *CYS* 基因,大麦中则发现了 13 个 *CYS* 基因^[5]。本研究通过全基因组生物信息学分析,从大豆中鉴定出 3 个亚家族共 19 个大豆 *CYS* 基因家族成员。基因家族成员数量分析表明大豆 *CYS* 基因的数量明显高于拟南芥、水稻等物种,表现出大豆的古四倍体特征^[16-17]。有报道介绍,水稻与拟南芥在进化过程中全基因组的复制仅仅发生过一次^[18-19]。而大豆则发生过 2 次复制,从而造成其基因组序列高度重复^[20]。大豆基因组的 2 次复制,导致了其基因多样化和诸多基因的丢失,并造成了很多染色体发生重排,使得大豆约 75% 的基因形成了多拷贝^[21-22]。

基因结构分析表明,大豆 *CYS* 基因家族各成员中,第一亚家族的 8 个基因均含有 1~3 个内含子;而其他 11 个属于第二和第三亚家族的 *CYS* 基因则根本没有内含子。利用蛋白编码基因起源的早期内含子模型加以解释^[23],该 11 个属于 Class II 和 Class III 的大豆 *CYS* 基因可能在进化中承受了较强的压力,从而造成了内含子丢失。蛋白序列结构域分析表明,大豆 *CYS* 家族基因编码蛋白中,大多数成员都只含有 1 个 Cystatin domain,但 *GmCYS12*、*GmCYS13* 和 *GmCYS15* 含有 2 个 Cystatin domain。根据现有报道,在拟南芥等物种中 *CYS* 蛋白所含 Cystatin domain 的数量与其催化活性之间并没有直接的关系^[24]。

现有的研究已经肯定植物体内的半胱氨酸蛋白酶抑制剂基因 *CYS* 与植物防御病虫害有重要关联^[3-6]。本研究通过对大豆各 *CYS* 基因的启动子分析发现,所有大豆 *CYS* 基因的上游调控序列中都含有大量与 ABA 或干旱诱导表达调控有关的 MYB、MYC 转录因子作用元件。此外,很多大豆 *CYS* 基因启动子序列中,还含有各种与环境因子响应和信号分子应答有关的顺式作用元件,如:ABRE 脱落酸响应元件、HSE 热激响应元件等。这可能意味着 *CYS*

基因家族成员在大豆生长发育过程中,不仅承担防御生物胁迫因子(如虫害、病害等)的作用,还同时肩负着应对部分非生物胁迫因子(如干旱胁迫、热胁迫等)侵害的作用。

UniGene 平台 EST 数据分析显示,14 个含有 EST 表达数据的大豆 *CYS* 基因,大多在子叶、花、叶片以及种皮等组织中表达。而害虫在取食农作物时,其消化酶主要为半胱氨酸蛋白酶。因此,大豆 *CYS* 基因在这些容易受到害虫取食的组织部位大量表达,将有助于增强植株对害虫的抵抗能力。由于 UniGene 数据库中的 EST 表达数据并没能包括所有的 19 个大豆 *CYS* 家族基因(*GmCYS1*、*GmCYS5*、*GmCYS7*、*GmCYS13* 以及 *GmCYS18* 未搜索到 EST 数据)。这可能是由于大豆 EST 数据库不全;或者是部分 *CYS* 基因只在特定的条件下才发生表达,因此难以被检测到;当然,也有可能是一些 *CYS* 基因在进化过程中发生了沉默,成为了假基因。

参考文献

- [1] Barrett A J, Fritz H, Grubb A, et al. Nomenclature and classification of the proteins homologous with the cysteine-proteinase inhibitor chicken cystatin[J]. Biochemical Journal, 1986, 236(1): 312.
- [2] Folta K, Wang H Y, Ricatirte S. Evaluation of strawberry cystatin gene family members as sting nematode antifeedants[J]. Hortscience, 2008, 43(4): 1155-1155.
- [3] Belenghi B, Acconcia F, Trovato M, et al. AtCYS1, a cystatin from *Arabidopsis thaliana*, suppresses hypersensitive cell death[J]. European Journal of Biochemistry, 2003, 270(12): 2593-2604.
- [4] Ohtsubo S, Taiyoji M, Kawase T, et al. Oryzacystatin-II, a cystatin from rice (*Oryza sativa* L. japonica), is a dimeric protein: Possible involvement of the interconversion between dimer and monomer in the regulation of the reactivity of Oryzacystatin-II[J]. Journal of Agricultural and Food Chemistry, 2007, 55(5): 1762-1766.
- [5] Martinez M, Cambra I, Carrillo L, et al. Characterization of the entire cystatin gene family in barley and their target cathepsin L-like cysteine-proteases, partners in the hordein mobilization during seed germination[J]. Plant Physiology, 2009, 151(3): 1531-1545.
- [6] Dou S J, Wei J, Cao Y H, et al. Molecular characterization and preliminary functional analysis of cystatin OC-I in rice[J]. Journal of Food Agriculture and Environment, 2011, 9(2): 235-239.
- [7] Huala E, Dickerman A W, Garcia-Hernandez M, et al. The *Arabidopsis* information resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant[J]. Nucleic Acids Research, 2001, 29(1): 102-105.
- [8] Ouyang S, Zhu W, Hamilton J, et al. The TIGR rice genome annotation resource: improvements and new features[J]. Nucleic Acids Research, 2007, 35(S1): 883-887.
- [9] Finn R D, Mistry J, Schuster-Bockler B, et al. Pfam: clans, web tools and services[J]. Nucleic Acids Research, 2006, 34: 247-251.

(下转第 666 页)

通过整合 Topo 载体和 LR 反应来构建双元载体 pGmARF16::mGmARF16, 主要是因为 GmARF16 基因组序列较长, 用限制性内切酶酶切连接很难获得目标载体。Topo 载体很方便进行 PCR 片段克隆, 只要在 PCR 正向引物 5' 端引入 CACC 就可以把片段克隆进去。pGWB433-M1 是专门开发出来并用 LR 重组反应来完成双元表达载体的构建, 此外, pGWB 系列载体具有多个表达元件组合, 为载体构建提供多种选择。

随着对大豆研究的深入进行, 大豆的遗传转化效率有了显著提供, 如子叶节法转化大豆最高可以达到 10% 左右, 这为大豆的基因工程改良提供又一重要平台^[14-15]。目前, 我们已通过农杆菌介导的子叶节法将该构建转化大豆, 获得了抗卡那霉素的再生绿苗, 进一步的试验正在进行之中。

参考文献

- [1] Leyser O. Dynamic integration of auxin transport and signaling[J]. Current Biology, 2006, 16:424-433.
- [2] Dharmasiri N, Estelle M. Auxin signaling and regulated protein degradation[J]. Trends Plant Science, 2004, 9:302-308.
- [3] Tiwari S B, Wang X J, Hagen G, et al. Aux/IAA proteins are active repressors and their stability and activity are modulated by auxin[J]. Plant Cell, 2001, 13:2809-2822.
- [4] Kepinski S, Leyser O. The *Arabidopsis* F-box protein TIR1 is an auxin receptor[J]. Nature, 2005, 435(7041):446-451.
- [5] Yang X Q, Lee S, So J H, et al. The IAA1 protein encoded by AXRS and is substrate of SCFTR1[J]. Plant Journal, 2004, 40:772-782.
- [6] Liacum E, Reed J W. Genetics of Aux/IAA and ARF action in plant growth and development[J]. Plant Molecular Biology, 2002, 49(3-4):387-400.
- [7] Tan X, Caideron-Villalobos L I, Sharon M, et al. Mechanism of auxin perception by the TIR1 ubiquitin ligase[J]. Nature, 2007, 446:640-645.
- [8] Reed J W. Roles and activities of Aux/IAA proteins in *Arabidopsis* [J]. Trends Plant Science, 2001, 6:420-425.
- [9] Nodine M D, Bartel D P. MicroRNAs prevent precocious gene expression and enable pattern formation during plant embryogenesis[J]. Genes & Development, 2010, 24(23):2678-2692.
- [10] Wang J W, Wang L J, Man Y B, et al. Control of root cap formation by microRNA-targeted auxin response factors in *Arabidopsis* [J]. Plant Cell, 2005, 17:2204-2216.
- [11] Liu P P, Montgomery T A, Fahlgren, et al. Repression of AUXIN RESPONSE FACTOR10 by microRNA160 is critical for seed germination and post-germination stages[J]. Plant Journal, 2007, 52(1):133-146.
- [12] Mallory A C, Barrel D P, Bartel B. MicroRNA-directed regulation of *Arabidopsis* AUXIN RESPONSE FACTOR17 is essential for proper development and modulates expression of early auxin response genes[J]. Plant Cell, 2005, 17:1360-1375.
- [13] Schmutz J, Cannon S B, Schlueter J. Genome sequence of the palaeopolyploid soybean[J]. Nature, 2010, 463:178-183.
- [14] Paz M M, Shou H, Guo Z, et al. Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant[J]. Euphytica, 2004, 136:167-179.
- [15] Paz M M, Martinez J C, Kalvig A B, et al. Improved cotyledonary node method using an alternative explant derived from mature seed for efficient *Agrobacterium*-mediated soybean transformation [J]. Plant Cell Report, 2006, 25:206-213.
- [10] Letunic I, Copley R R, Pils B, et al. SMART 5: domains in the context of genomes and networks[J]. Nucleic Acids Research, 2006, 34:257-260.
- [11] Yang S H, Zhang X H, Yue J X, et al. Recent duplications dominate NBS-encoding gene expansion in two woody species[J]. Molecular Genetics and Genomics, 2008, 280(3):187-198.
- [12] 郭安源, 朱其惠, 陈新, 等. GSDS: 基因结构显示系统[J]. 遗传, 2007, 29(8):1023-1026. (Guo A Y, Zhu Q H, Chen X, et al. GSDS: a gene structure display server [J]. Hereditas, 2007, 29(8):1023-1026.)
- [13] 刘仁虎, 孟金陵. MapDraw, 在 Excel 中绘制遗传连锁图的宏[J]. 遗传, 2003, 25(3):317-321. (Liu R H, Meng J L. MapDraw: a microsoft excel macro for drawing genetic linkage maps based on given genetic linkage data[J]. Hereditas, 2003, 25(3):317-321.)
- [14] Tamura K, Peterson D, Peterson N, et al. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods[J]. Molecular Biology and Evolution, 2011, 28(10):2731-2739.
- [15] Jiang M F, Li S W, Chen M, et al. Molecular cloning and expression of cDNA encoding the cysteine proteinase inhibitor from upland cotton[J]. Journal of Plant Biology, 2009, 52(5):426-432.
- [16] Mukherjee K, Brocchieri L, Burglin T R. A comprehensive classification and evolutionary analysis of plant homeobox genes[J]. Molecular Biology and Evolution, 2009, 26(12):2775-2794.
- [17] Hay A, Tsiantis M. KNOX genes: versatile regulators of plant development and diversity [J]. Development, 2010, 137(19):3153-3165.
- [18] Blanc G, Barakat A, Guyot R, et al. Extensive duplication and reshuffling in the *Arabidopsis* genome[J]. Plant Cell, 2000, 12(7):1093-1101.
- [19] Wang X Y, Shi X L, Hao B L, et al. Duplication and DNA segmental loss in the rice genome: implications for diploidization[J]. New Phytologist, 2005, 165(3):937-946.
- [20] Schlueter J A, Dixon P, Granger C, et al. Mining EST databases to resolve evolutionary events in major crop species [J]. Genome, 2004, 47(5):868-876.
- [21] Schmutz J, Cannon S B, Schlueter J, et al. Genome sequence of the palaeopolyploid soybean[J]. Nature, 2010, 463:178-183.
- [22] Severin A J, Cannon S B, Graham M M, et al. Changes in twelve homeologous genomic regions in soybean following three rounds of polyploidy[J]. Plant Cell, 2011, 23(9):3129-3136.
- [23] Zhaxybayeva O, Gogarten J P. Spliceosomal introns: New insights into their evolution[J]. Current Biology, 2003, 13(19):764-766.
- [24] Heeg C, Kruse C, Jost R, et al. Analysis of the *Arabidopsis* O-acetylserine(thiol) lyase gene family demonstrates compartment-specific differences in the regulation of cysteine synthesis[J]. Plant Cell, 2008, 20(1):168-185.

(上接第 660 页)