

研究报告

Research Report

大豆 GST 基因家族全基因组筛选、分类和表达

江董丽¹✉, 才华²✉, 端木慧子²✉, 朱延明²✉

1 塔里木大学生命科学学院, 阿拉尔, 843300

2 东北农业大学生命科学学院, 哈尔滨, 150030

✉ 通讯作者, ymzhu2001@neau.edu.cn; ✉ 作者

豆科基因组学与遗传学, 2013年, 第4卷, 第2篇 doi: 10.5376/lgg.cn.2013.04.0002

收稿日期: 2013年03月18日 接受日期: 2013年03月18日 发表日期: 2013年03月18日

© 2013 BioPublisher 生命科学中文期刊出版平台

本文首次发表在《分子植物育种》2013年, 第11卷, 第5期上。现依据版权所有人授权的许可协议, 采用 [Creative Commons Attribution License](#), 协议对其进行授权, 再次发表与传播。只要对原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

摘要 利用生物信息学手段, 结合公共大豆基因组数据库和大豆发育表达芯片数据获得大豆 GST 家族基因序列、蛋白序列和染色体位置等信息并进一步对基因的组织表达等进行分析。结果显示大豆中含 94 个 GST 家族基因, 根据系统发育分析将这些 GST 基因分成 5 个亚家族; 定位分析表明, 94 个 GST 基因分布于大豆的 16 条染色体上。表达分析结果表明, 14 个不同发育阶段, 大多成员至少在一个组织中表达, 11 差异表达的基因中有 7 在根中表达, 另外 4 在其它部位优势表达, 基因表达具有一定特异性。本研究为进一步研究 GST 家族的功能及其抗逆利用提供基础。

关键词 GST 家族; 全基因组筛选; 分类分析; 差异表达基因

Genome-Wide Filter, Classification and Expression Analysis of GST Gene Family in Soybean

Jiang Dongli¹✉, Cai Hua²✉, Duanmu Huizi²✉, Zhu Yanming²✉

1 College of Life Sciences, Tarim University, Alar, 843300

2 Key Laboratory of Agricultural Biological Functional Gene, Northeast Agricultural University, Harbin, 150030

✉ Corresponding author, ymzhu2001@neau.edu.cn; ✉ Authors

Abstract Based on bioinformatic methods, soybean genome database and expression microarray data, soybean GST genes family were identified in this study. And their gene sequences, protein sequences and chromosome location were obtained. All of ninety-four GST genes from soybean were classified into 5 subfamilies according to phylogenetic relationship. And they were distributed on 16 chromosomes. Based on the soybean developmental expression database, the expression pattern of the GST gene family at different development organs and stages was analyzed. Eleven differentially expressed genes were found, and seven of them were highly expressed in root at each developmental stage by microarray. They were differentially expressed in different tissues and developmental stages. These results are helpful for their functional analysis and using for crop anti-stress improvement.

Keywords GST family, Genome-wide screening, Classify analysis, Differentially expressed gene

谷胱甘肽 S-转移酶(GSTs)是一类普遍存在的酶, 这类酶用于催化谷胱甘肽与毒性生物异源物质或氧化产物结合, 从而促进此类物质的代谢、区域化隔离或清除(Wilce and Parker, 1994; 王臻昱等, 2012)。经典的反应模式为, GSTs 催化谷胱甘肽(γ -谷氨酰-半胱氨酰-甘氨酸; GSH)转移到含有活性

亲电中心的辅酶(R-X)上建立一个极性的 S-谷胱甘肽反应产物(R-SG)。谷胱甘肽 S-转移酶是一个庞大而多样的基因家族, 根据氨基酸序列相似性将其分为 tau、phi、zeta、theta、lambda 和 DHAR (谷胱甘肽依赖的抗坏血酸还原酶) 6 个亚类, 其中 tau 和 phi 两类 GST 在植物中含量最为丰富(Dixon et al., 2002;

Jain et al., 2010)。GST 基因序列同源性虽然较低(约为 25%),但通过对大量 GSTs 蛋白结构的研究发现(McGonigle et al., 2000; Thom et al., 2001),这类蛋白具有高度保守的结构特征。

GST 蛋白在多个代谢反应中起到重要作用。Edwards 等(2000)发现玉米中 GST 蛋白能够催化谷胱甘肽与氯均三嗪阿特拉津结合,从而保护植物免受这类除草剂的伤害。Dalton 等(2009)发现大豆根瘤中含有丰富的 GSTs,研究表明根中 GSTs 对大豆根瘤的生物固氮有促进作用。除此之外,GST 家族蛋白,也在植物抗击非生物胁迫反应中发挥积极作用。研究表明,谷胱甘肽 S-转移酶(GSTs)可通过降解胁迫条件下细胞内积累的毒性氧化物,从而保护植物免受氧化伤害(Sheehan et al., 2001; 王臻昱等, 2012)。已有大量研究表明 GST 转基因植物能够增强植物对逆境,如低温(Huang et al., 2009; 赵风云等, 2006)、干旱(George et al., 2010)、盐碱(Ji et al., 2010)及重金属(Moons, 2003; 高可辉等, 2011)等的耐受能力。

大豆作为一种重要的农作物,对于提高其抗逆性改良品种有很高价值。虽然对于 GSTs 基因已有很多报道,但是在大豆对非生物逆境胁迫反应,具体是哪一个或是哪一类 GSTs 基因起着关键作用还不清楚。大豆基因组全序列信息和相关信息的注释和公布(http://www.phytozome.net/search.php?show=text&method=Org_Osativa/) (Schmutz et al., 2010),使得在全基因组水平上运用生物信息学方法系统地分析大豆 GST 基因家族成为可能。本研究利用大豆全基因组数据和大豆基因芯片数据(<http://www.soybase.org/>)用生物信息学方法,筛选所有的 GST 基因家族,了该家族基因的染色体定位、分类、系统发育、启动子区的序列特征与转录谱变化值,为 GST 基因家族成员的功能研究及利用其提高农作物耐逆性提供参考。

1 结果与分析

1.1 大豆 GST 基因家族的筛选

按照 Phytozome 大豆基因组数据库数据中的注释信息,并结合 Pfam 结构域分析,初步筛选出 118 个候选大豆 GST 基因,其中 9 个基因因不含完整开放读码框被排除; SMART 结构域分析剔除了 15 个

不含有或含有不完整 GST 结构域(基因),并结合 McGonigle 等(2000)对大豆 GST 家族的综合分析结果后最终得到 94 个大豆 GST 基因。大豆 GST 基因家族 94 个基因中蛋白等电点范围从 4.75 到 10.22 不等,其蛋白最短的有 53 个氨基酸残基,最长的有 401 个,跨度较大。

1.2 大豆 GST 基因家族的序列和结构域分析

分析基因结构大豆 GST 家族基因,其具有的内含子较少(图 1), 47 个基因(约 50.00%)含有 1 个内含子, 26 个基因(27.66%)含有 2 个内含子,只有 4 个基因(0.04%)不含有内含子,剩下 18.10%的基因含有 3~9 个内含子。

利用 Pfam 和 SMART 在线工具分析结构域,结果表明(表 1),大豆 GST 家族蛋白至少有一个 GST 结构域(GST_N 或 GST_C),有的存在两个结构域(GS-T_N 和 GST_C),其中 25 个家族成员含有 1 个结构域(GST_N 或 GST_C),其它成员均含有 2 个结构域。

1.3 大豆 GST 基因家族的系统发育分析与分类

根据大豆 GST 基因家族的系统发育树(图 1),将大豆分为五个亚家族,即 Tau、Phi、Lambda、Zeta 和 Theta 类(图 1)。对 94 个大豆 GST 家族蛋白和拟南芥 GST 家族蛋白,进行多重序列联配(ClustalX2 软件)和系统进化分析(MEGA5.1 软件),结果显示拟南芥 GST 基因归属于除了 Lambda 亚家族类的其他四个亚家族(结果未显示)。

1.4 大豆 GST 基因家族在染色体上的分布

根据 GST 家族的基因染色体定位信息,获得 94 个 GST 基因在大豆染色体上的分布图(MapInspect 软件)(图 2)。这些基因分布在 16 条大豆染色体上,其中第 4、第 9、第 12 和第 16 号染色体上不含有 GST 基因,第 7 号染色体还有的成员数量最多,有 13 个,第 2、第 8、第 15 和第 18 号染色体分布的成员数量较多,有 9~11 个,第 19 号染色体含有的成员数量最少,只有 1 个,其它各染色体含有 3~6 个 GST 基因。

将这 94 个大豆 GST 家族基因一一比对(DNAMAN 软件),并结合旁系同源基因的概念(Heikoff et al., 1997)和同源基因的判断标准(Yang et al., 2008),发现了 12 对旁系同源基因(在图 2 中用

表 1 大豆 GST 基因的结构域及其启动子中 ABRE, DRE, ERE, LTRE, MYB 和 MYC 元件的数量分布
Table 1 Domain in GST genes and distribution of ABRE, DRE, ERE, LTRE, MYB, and MYC cis-acting elements in soybean GST gene promoters

基因序列号 Gene ID number	结构域(Pfam) Domain (Pfam)	ABR	DRE	ERE	LTRE	MYB	MYC
Glyma01g04690	GST-N(02798), GST-C(00043)	2	0	1	1	2	28
Glyma01g04700	GST-N, GST-C	6	0	2	2	17	22
Glyma01g04710	GST-N, GST-C	1	0	3	0	18	16
Glyma01g26220	GST-N, GST-C	0	0	0	1	11	8
Glyma01g26230	GST-N, GST-C	5	0	3	0	7	8
Glyma01g34360	GST-N	1	0	0	0	17	20
Glyma02g02860	GST-N, GST-C	1	1	1	1	11	8
Glyma02g02870	GST-N	2	0	3	0	21	12
Glyma02g02880	GST-N, GST-C	0	1	2	2	22	14
Glyma02g11045	GST-N	8	2	0	3	12	34
Glyma02g17330	GST-N, GST-C	0	0	2	1	15	18
Glyma02g17340	GST-N, GST-C	3	0	0	2	21	22
Glyma02g33780	GST-N, GST-C	4	0	0	4	12	8
Glyma02g40760	GST-C	3	2	0	2	26	16
Glyma02g45328	GST-N, GST-C	5	0	1	0	24	34
Glyma02g45336	GST-N, GST-C	2	0	1	0	14	18
Glyma03g16563	GST-N, GST-C	2	0	1	0	15	12
Glyma03g16600	GST-N, GST-C	0	0	1	1	10	10
Glyma03g33340	GST-C	0	4	3	5	8	16
Glyma05g29370	GST-N, GST-C	1	2	1	6	20	10
Glyma05g29381	GST-N, GST-C	4	0	0	2	17	10
Glyma05g29390	GST-N, GST-C	1	1	0	2	10	12
Glyma05g29400	GST-N, GST-C	2	0	0	0	9	23
Glyma05g37905	GST-N	2	0	1	0	20	12
Glyma05g37910	GST-N	3	1	0	1	9	14
Glyma06g10390	GST-N	10	2	2	2	9	20
Glyma06g12400	GST-N, GST-C	0	0	3	0	18	20
Glyma06g20730	GST-N	0	0	0	0	10	22
Glyma07g08202	GST-C	1	2	2	1	19	12
Glyma07g16800	GST-N, GST-C	1	0	2	1	10	18
Glyma07g16810	GST-N, GST-C	3	0	2	2	17	20
Glyma07g16830	GST-N, GST-C	8	0	1	0	13	26
Glyma07g16840	GST-N, GST-C	2	0	0	0	6	4
Glyma07g16850	GST-N, GST-C	1	0	0	0	12	6
Glyma07g16855	GST-N, GST-C	4	0	1	1	15	12
Glyma07g16860	GST-N, GST-C	3	0	1	0	15	8
Glyma07g16876	GST-N	5	0	3	2	11	8
Glyma07g16910	GST-N, GST-C	2	2	2	2	5	6
Glyma07g16925	GST-N, GST-C	2	0	5	0	9	14
Glyma07g16940	GST-N, GST-C	3	0	1	0	4	4
Glyma07g34820	GST-C	0	0	1	0	15	8
Glyma08g01691	GST-N	3	1	1	2	21	18
Glyma08g12510	GST-N, GST-C	3	0	0	1	14	16
Glyma08g12520	GST-N, GST-C	2	1	1	1	15	22
Glyma08g12530	GST-N, GST-C	0	0	1	1	14	24
Glyma08g18640	GST-N, GST-C	1	0	0	2	20	26
Glyma08g18660	GST-N, GST-C	2	0	2	0	18	10

续表1
Continuing table 1

基因序列号 Gene ID number	结构域(Pfam) Domain (Pfam)	ABR	DRE	ERE	LTRE	MYB	MYC
Glyma08g18671	GST-N, GST-C	1	0	0	6	18	26
Glyma08g18680	GST-N, GST-C	0	0	0	0	23	28
Glyma08g18690	GST-N, GST-C	2	0	1	0	5	8
Glyma08g41960	GST-N, GST-C	4	0	2	0	13	36
Glyma08g41970	GST-N, GST-C	3	0	2	0	16	8
Glyma10g02460	GST-N, GST-C	0	0	4	0	5	20
Glyma10g04570	GST-N, GST-C	7	0	1	1	7	16
Glyma10g05480	GST-C	0	0	1	2	16	6
Glyma10g33650	GST-N, GST-C	5	0	2	0	13	14
Glyma10g33691	GST-C	1	0	1	0	11	6
Glyma11g31330	GST-N, GST-C	7	3	2	0	15	18
Glyma11g33260	GST-N, GST-C	11	0	1	4	17	4
Glyma13g09650	GST-N, GST-C	6	0	1	1	13	6
Glyma13g15550	GST-C	3	0	1	2	10	14
Glyma13g19130	GST-N, GST-C	2	0	2	1	11	10
Glyma13g19830	GST-C	4	0	0	0	9	14
Glyma13g19840	GST-C	3	3	3	2	12	10
Glyma14g03461	GST-C	1	0	1	0	25	28
Glyma14g03470	GST-N, GST-C	12	0	1	0	20	30
Glyma14g24620	GST-C	2	0	0	1	11	6
Glyma14g31915	GST-C	4	2	0	1	25	12
Glyma14g39090	GST-N, GST-C	3	0	0	1	27	28
Glyma15g40190	GST-N, GST-C	1	0	1	0	7	10
Glyma15g40200	GST-N, GST-C	26	1	2	1	8	24
Glyma15g40220	GST-N, GST-C	0	0	1	0	19	24
Glyma15g40240	GST-N, GST-C	0	0	1	0	9	12
Glyma15g40250	GST-N, GST-C	1	2	1	1	30	16
Glyma15g40260	GST-N, GST-C	6	0	2	0	14	30
Glyma15g40280	GST-N	1	0	0	0	27	20
Glyma15g40290	GST-N, GST-C	5	0	1	0	16	16
Glyma15g40305	GST-C	2	0	3	0	12	10
Glyma17g00691	GST-N, GST-C	2	0	1	0	11	10
Glyma17g00700	GST-N, GST-C	0	1	1	1	12	12
Glyma17g04680	GST-N, GST-C	1	2	4	2	10	4
Glyma18g04960	GST-N, GST-C	5	0	1	0	17	20
Glyma18g05820	GST-N, GST-C	3	0	0	2	9	10
Glyma18g13620	GST-N, GST-C	0	0	1	0	14	12
Glyma18g13630	GST-N, GST-C	0	2	2	3	7	10
Glyma18g13640	GST-N, GST-C	1	3	0	3	15	6
Glyma18g16850	GST-N, GST-C	0	0	0	0	18	24
Glyma18g41340	GST-N, GST-C	1	1	4	0	6	24
Glyma18g41350	GST-N, GST-C	0	0	2	0	15	10
Glyma18g41410	GST-N, GST-C	4	1	2	1	8	16
Glyma19g36080	GST-C	4	1	0	0	23	16
Glyma20g02490	GST-N, GST-C	1	0	4	3	21	8
Glyma20g23420	GST-N, GST-C	1	0	3	1	9	18
Glyma20g33951	GST-N	7	2	0	3	7	20

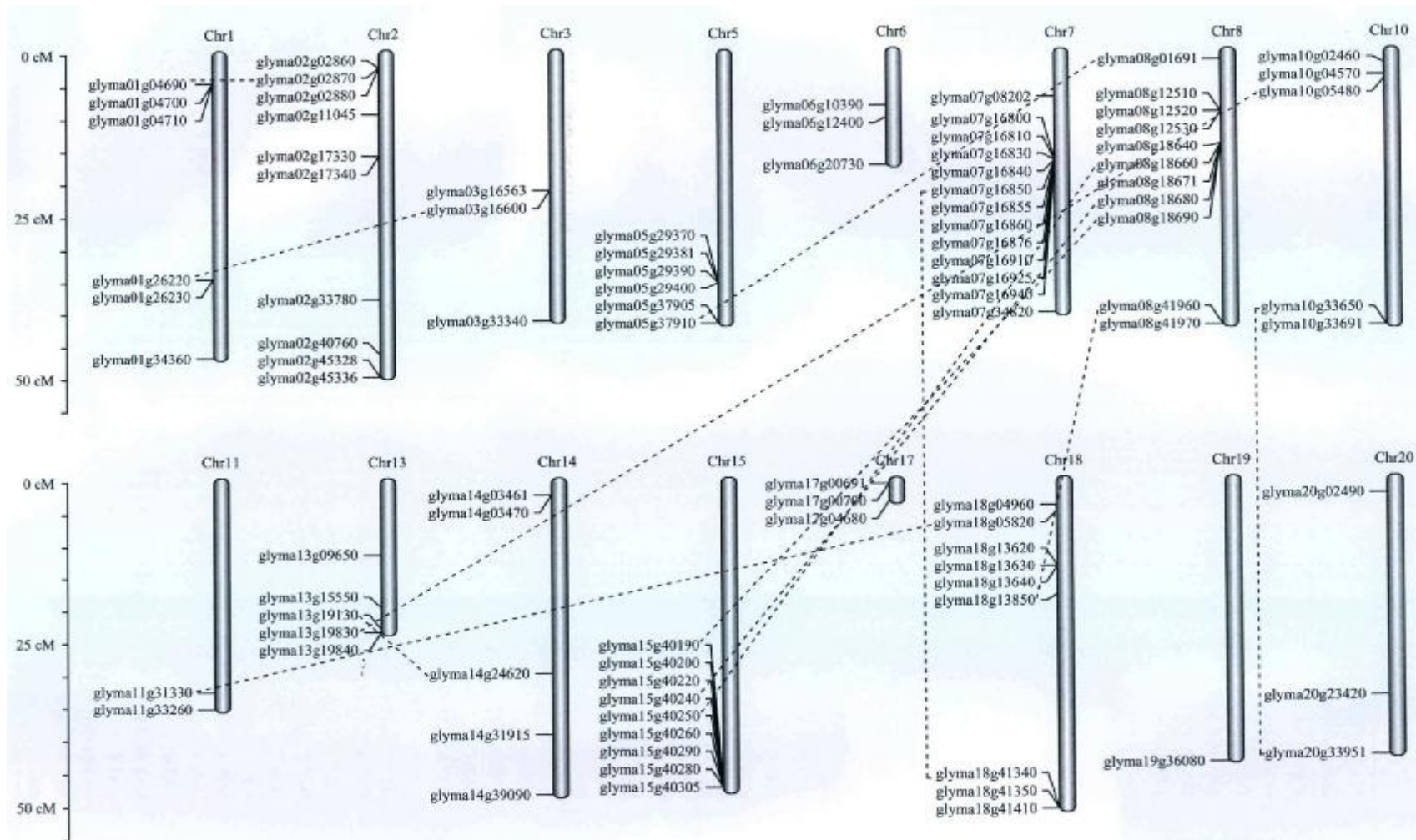


图 2 大豆 GST 基因的染色体分布

注:基因组进化中旁系同源基因重复事件用直线连接表示

Figure 2 Chromosome distribution of GST genes in soybean genome

Note: Genes related by endo-reduplication events during genome evolution (homologous genes) are connected by line

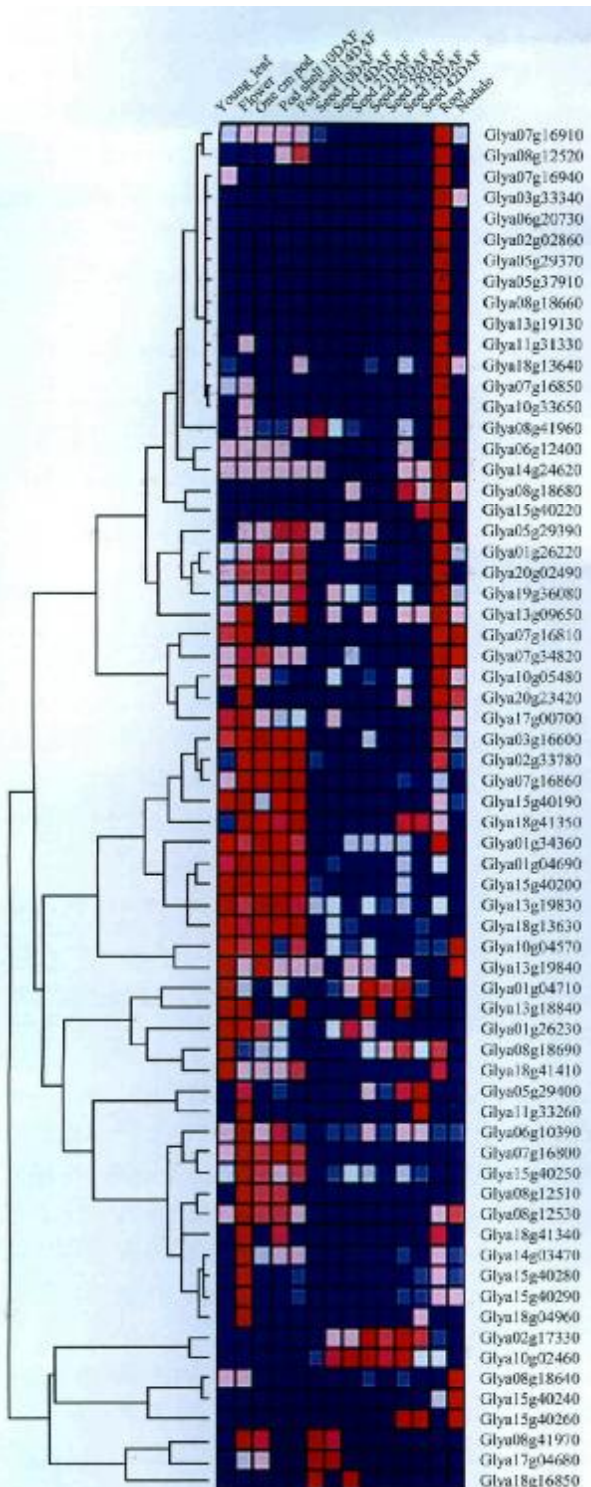


图 3 大豆 GST 基因在不同组织和生长发育阶段的表达
注: 方块内颜色显示芯片表达热度: 红色最高, 其次为粉红, 白色, 浅蓝, 深蓝, 紫色, 其中紫色为 0
Figure 3 Expression profile of soybean GST genes in different tissues and developmental stages
Note: Colors in square presented heatmap scale of expression in chips: red is the highest, pink is higher, then is white, light blue, dark blue, and purple is zero

Glyma08g18640 (373 TPM)和 Glyma07g16910 (369 TPM), 其余基因的表达量都在 300 TPM 以下, 最少的 Glyma02g02860 表达量只有 1 TPM。大豆 GST 基因中有 5 个基因只在根中表达, 包括 Glyma02g02860 (Tau 类)、Glyma05g29370 (Tau 类)、Glyma05g37910 (Theta 类)、Gm08g18660 (Tau 类)和 Glyma13g19130 (Tau 类)。另外, 有 11 个差异表达基因, 其中, Glyma08g18640 (Tau 类)在根瘤中优势表达, Glyma14g03470 (Phi 类)、Glyma15g40280 (Tau 类)和 Glyma15g40290 (Tau 类)这 3 个基因在花中优势表达, Glyma03g33340 (Lambda 类)、Glyma06g20730 (Tau 类)、Glyma07g16850 (Tau 类)、Glyma07g16910 (Tau 类)、Glyma08g12520 (Tau 类)、Glyma10g33650 (Tau 类)和 Glyma18g13640(Phi 类)这 7 个基因在根中的表达占有优势。

2 讨论

分析大豆全基因组后, 本研究共扫描筛选出 94 个大豆 GST 家族基因, 分属于 5 个亚类(图 1)。已知拟南芥包括 48 个 GST 基因(Wagner et al., 2002), 其中 Tau 和 Phi 类 GSTs 数量最多, 分别是 28 和 13 个基因, 而 Theta 类 GSTs 只有 3 个, Zeta 和 Lambda 类 GSTs 分别只有 2 个。而由图 1 可知, 大豆有 63 个 Tau 类, 17 个 Phi 类, 8 个 Lambda 类, 3 个 Zeta 类和 3 个 Theta 类 GSTs。大豆 Tau 和 Zeta 类 GSTs 的丰度分布都与拟南芥类似, 而大豆的 Phi 类 GSTs 比例相对小于拟南芥, 大豆 Lambda 类 GATs 比例大于拟南芥。大豆中 GST 基因家族数目近乎于拟南芥的两倍, 而分类却与拟南芥一样。

对启动子区的顺式作用元件分析中(表 1), 所有 GST 基因家族都含有 MYC 和 MYB 这两种顺式作用元件, 大部分家族成员含有 ABRE、DRE、ERE 和 LTRE 等顺式作用元件。外界非生物胁迫环境刺激应答调控的主要途径是 ABA 依赖型和非依赖型(向旭和傅家瑞, 1998; 龙海涛等, 2004)。在 ABA 依赖型和非依赖型调控机制中, ABRE、DRE、MYB 和 MYC 等元件起重要的作用。在 ABA 依赖型中 ABA 诱导基因的顺式作用元件有 ABRE、MYB、MYC 等, 该信号系统通过 ABA 合成 MYB、MYC 转录因子启动下游抗逆基因的表达(Abe et al., 1997; 2003; Agarwal et al., 2006)。在 ABA 非依赖型信号调控中, 传感蛋白被信号激活后向 DREB 传递, 再结合 DRE 顺式作用元件, 下游抗逆基因即可被启动表达(Shinozaki and Yamaguchi-Shinozaki, 2000; 李乐等, 2011; 郑炜君等, 2012)。同时, 在植物逆境胁迫中, ABRE、DRE 和 LTRE 等顺式作用元件的作用也较重要。其中, ABRE 诱导大量特异性蛋白, 在 ABA 应答中, 产生一些植物内调节作用¹²(Nakashima et al., 2006; 李

乐等, 2011); DRE 顺式作用元件结合 DREB 转录因子后在干旱、低温和高盐条件下调节下游一系列逆境应答基因的表达(Dubouzet et al., 2003; Qin et al., 2004; DiazMartin et al., 2005), LTRE 对 ABA 应答调节冷和干旱等逆境(Baker et al., 1994; Dunn et al., 1998)。对康乃馨 GST1 基因的研究发现 ERE, 即乙烯应答元件, 在花瓣衰老调节中起作用(Itzhaki et al., 1994)。有研究, 在应答外界非生物胁迫时, ABA 依赖型与 ABA 非依赖型这 2 种调控机制相互交叉形成调控网络作用, 而非彼此独立(Shinozaki and YamaguchiShinozaki, 2000)。此外, 还发现了三个不含有 ABRE、DRE、ERE 和 LTRE 的基因, 说明 GST 基因家族可能还存在有其它的调控机制。从表 1 中分析发现 ABRE、ERE、LTRE、MYB 和 MYC 等顺式作用元件在 29 个 GST 基因的启动子区都存在, 其中在 12 个基因的启动子区中还含有 DRE 顺式作用元件。启动子分析指出大豆中 GST 基因家族基因在应答外界的非生物胁迫时可能是通过 ABA 依赖与非依赖型调控, 或是两者相互交联采用更精细的方式起到调控作用(郑炜君等, 2012)。

基因差异表达分析利用基因芯片表达数据进行, 一共有 11 个 GST 基因差异表达。在根中特异表达的有 7 个基因, 其中, Tau 类有 5 个, Phi 类和 Lambda 类分别有 1 个, 说明 GST 基因家族中至少 3 个亚家族对植物的根的发育有影响, 且 Tau 类为主; 另外在根瘤、花等部位优势表达的有 4 个基因中就有 3 个是 Tau 类, 1 个是 Phi 类, 从而显示出该家族的基因表达具有多样性。但由于该表达数据库并没有包含所有 GST 基因家族的表达谱信息, 所以分析结果可能具有一定的片面性, 不全面。

对大豆全基因组数据(JGI Glyma1.1 annotation)进行鉴定与分析, 一共筛选得到 94 个大豆 GST 家族基因成员, 可将其分为 5 个亚家族(Tau, Phi, Lambda, Theta 和 Zeta 类), 在 16 条染色体上均有分布, 均至少有两个逆境反应顺式作用元件存在其启动子区中, 其转录谱表达分析显示其具有一定的组织特异性。

大豆 GST 基因在不同组织和生长发育阶段的表达强度分析, 大多数 GST 基因在根部显著表达(图 3)。66 个具有表达值的基因中只有 16 个在根中不表达, 这可能是由于根是植物的重要水分和矿物质营养的转运和吸收器官, 在盐和干旱等胁迫后, 根成为识别和受伤害的最主要的器官, 能对逆境胁迫作出快速应答。已有文献证实一些根部 GST 基因在甘薯、桤柳等植物的胁迫耐受中诱导表达并行使功能(刘珣等, 2009; 杨桂燕等, 2011, <http://www.paper.edu.cn/releasepaper/content/201110-207/>), 因此可以进一步推测大豆 GST 基因与逆境胁迫有关。尤其是 Glyma02g02860 (Tau 类)、

Glyma05g29370 (Tau 类)、Glyma05g37910 (Theta 类)、Glyma08g18660 (Tau 类)和 Glyma13g19130 (Tau 类)这 5 个基因仅在根中表达, 暗示其有可能在遭受胁迫时起关键作用。另外, 有实验表明 GST 基因与植物对除草剂的抗性相关(Karavangeli et al, 2005; 朱其松等, 2009)。有研究表明, 除草剂乙草胺和阿特拉津对不同组织的诱导作用各不相同, 乙草胺诱导玉米叶中 GSTs 活性小幅增加, 抑制根中 GSTs 活性; 阿特拉津诱导玉米根中 GSTs 活性增加, 抑制叶中 GSTs 的活性(郭玉莲等, 2008)。Pascal 等(2000)发现除草剂 acifluorfen 诱导的小麦 GST 主要在幼苗的地上组织表达, 因此本研究中显示的在幼叶中高表达的 Glyma01g04710 和 Glyma13g19840 等基因可能在大豆对除草剂耐受上起重要作用。同时, 本研究发现大多数 GST 基因在花中同样有表达, 推测 Glyma14g03470 (Phi 类)、Glyma15g40280 (Tau 类)和 Glyma15g40290 (Tau 类)这些在花中显著表达的基因可能参与大豆开花期的生长发育过程。总之, 本实验通过结合多种数据资源, 对大豆 GST 基因家族进行了系统的分析, 并根据 GST 基因表达的组织特异性筛选出一些耐逆的候选基因, 为大豆的品种改良和提高大豆对非生物胁迫的耐受性提供了理论基础。

3 材料与方法

3.1 材料

从 Phytozome 网站 (<http://www.phytozome.net/index.php/>) 大豆最新数据库版本 JGI Glyma1.1 annotation 中下载大豆(Glycine max)全基因组数据、相关信息和 cDNA 数据。从 TAIR (Huala et al., 2001)网站(<http://arabidopsis.org/>)下载拟南芥 GST 家族的基因序列和蛋白序列。从 Soybase 网站 (<http://www.soybase.org/>) 下载大豆基因芯片数据。

3.2 GST 基因家族成员筛选、结构域分析和基因结构分析

Phytozome 网站中大豆基因组数据库(<http://www.phytozome.net/index.php/>)上的序列已经经过结构域等注释, 直接搜索“glutathione S-transferas”即可获得具有 GST 结构域和预测的 GST 基因家族, 下载获得每个基因序列的相关信息。使用 Pfam 数据库工具(Finn et al., 2010), 同时利用已经鉴定出的拟南芥中的 GST 基因家族成员的相关信息, 获得其典型的结构域。再使用 SMART 在线网址 (<http://smart.emblheidelberg.de/>) (Letunic et al., 2012) 对其结构域进行进一步的分析, 去除无典型的或不完整的 GST 结构域序列, 并结合 McGonigle 等(2000)对大豆 GST 家族的综合分析结果, 最终筛选获得

大豆 GST 基因家族成员。这些基因的外显子-内含子结构图运用 GSDS (郭安源等, 2007) 在线网址 (<http://gsds.cbi.pku.edu.cn/>) 绘制。

3.3 GST 蛋白多序列比对和系统进化树的构建

多序列比对分析大豆和拟南芥 GST 蛋白利用的是 Clustal X2 程序, 得到的序列比对结果使用 MEGA5.1 (<http://megasoftware.net/>) 下载程序采用邻接法(bootstrap 值设置为 500)生成 GST 基因家族的系统进化树(Tamura et al., 2011)。

3.4 GST 家族基因的染色体定位

结合 JGI Glyma 1.1 Annotation (大豆基因组最新数据版本) 中基因在染色体中的信息, 利用 MapInspect 本地软件, 获得 94 个 GST 家族基因在染色体上的位置图。大豆 GST 家族基因中的重复片段分析利用 DNAMAN 软件, 采用 Yang 等(2008)与 Heikoff 等(1997) 文献中提到的方法, 判定其旁系同源基因, 即在不同染色体上的 2 个基因的序列比对部分的长度要涵盖其较长序列的 80%, 并且其序列比对部分的相似性要大于 70%, 然后在染色体上标注分析结果。

3.5 GST 家族基因的启动子区分析

启动子区设为大豆 GST 基因起始密码子上游 2 000 bp 的序列, 从 Phytozome 大豆基因组数据库数据中获得其序列, 主要通过 PLACE 30.0(植物 DNA 顺式作用调控元件数据库)在线网站, 并结合 PlantCARE (植物顺式作用调控元件数据库)在线网站(郑伟君等, 2012), 分析这个家族成员启动子区的顺式作用元件。

3.6 GST 基因的转录谱表达

从网站(<http://www.soybase.org/>) 获得这 94 个基因在 14 个不同器官和生长发育阶段的基因芯片数据, 其差异表达基因的分析利用 Jain 等(2007)的方法。运用在线网址 (<http://genepattern.broadinstitute.org/gp/pages/index.jsf?lsid=HierarchicalClustering/>) 对芯片表达数据进行数据聚类分析。

作者贡献

才华和江董丽是本研究的实验设计和实验研究的执行人; 江董丽完成数据分析, 论文初稿的写作; 端木慧子参与试验结果分析和论文终稿的修改; 朱延明和才华是项目的构思者及负责人, 指导实验设计、数据分析, 论文写作与修改。全体作者都阅读并同意最终的文本。

致谢

本研究由国家自然科学基金项目(31171578)、国家转基因生物新品种培育科技重大专项(2011ZX08004002-002)和黑龙江省高校科技创新团队建设计划(2011TD005)共同资助。

参考文献

- Abe H., Yamaguchi-Shinozaki K., Urao T., Iwasaki T., Hosokawa D., and Shinozaki K., 1997, Role of Arabidopsis MYC and MYB Homologs in drought and abscisic acid-regulated gene expression, *Plant Cell*, 9(10): 1859-1868
- Abe H., Urao T., Ito T., Seki M., Shinozaki K., and Yamaguchi-Shinozaki K., 2003, Arabidopsis AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling, *Plant Cell*, 15(1): 63-78
- Agarwal M., Hao Y., Kapoor A., Dong C.H., Fujii H., Zheng X., and Zhu J.K., 2006, A R2R3 type MYB transcription factor is involved in the cold regulation of CBF genes and in acquired freezing tolerance, *J. Biol. Chem.*, 281(49): 37636-37645
- Baker S.S., Wilhelm K.S., and Thomashow M.F., 1994, The 5'-region of Arabidopsis thaliana cor15a has cis-acting elements that confer cold-, drought and ABA-regulated gene expression, *Plant Mol. Biol.*, 24(5): 701-713
- Dalton D.A., Boniface C., Turner Z., Lindahl A., Kim H. J., Jelinek L., Govindarajulu M., Finger R.E., and Taylor C.G., 2009, Physiological roles of glutathione S-transferases in soybean root nodules, *Plant Physiol.*, 150(1): 521-530
- Diaz-Martin J., Almoguera C., Prieto-Dapena P., Espinosa J.M., and Jordano J., 2005, Functional interaction between two transcription factors involved in the developmental regulation of a small heat stress protein gene promoter, *Plant Physiol.*, 139(3): 1483-1494
- Dixon D.P., Laphorn A., and Edwards R., 2002, Plant glutathione transferases, *Genome Biol.*, 3(3): REVIEWS3004
- Dubouzet J.G., Sakuma Y., Ito Y., Kasuga M., Dubouzet E.G., Miura S., Seki M., Shinozaki K., and Yamaguchi-Shinozaki K., 2003, OsDREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt and cold-responsive gene expression, *Plant J.*, 33(4): 751-763
- Dunn M.A., White A.J., Vural S., and Hughes M.A., 1998, Identification of promoter elements in a low-temperature-responsive gene (b1t4.9) from barley (*Hordeum vulgare* L.), *Plant Mol. Biol.*, 38(4): 551-564
- Edwards R., Dixon D.P., and Walbot V., 2000, Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health, *Trends Plant Sci.*, 5(5): 193-198
- Finn R.D., Mistry J., Tate J., Coghill P., Heger A., Pollington J.E., Gavin O.L., Gunasekaran P., Ceric G., Forslund K., Holm L., Sonnhammer E.L., Eddy S.R., and Bateman A., 2010, The Pfam protein families database, *Nucleic Acids Res.*, 38: 211-222
- Guo A.Y., Zhu Q.H., Chen X., and Luo J.C., 2007, GSDS: a gene structure display server, *Yichuan (Hereditas (Beijing))*, 29(8): 1023-1026 (郭安源, 朱其惠, 陈新, 罗静初, 2007, GSDS: 基因结构显示系统, *遗传*, 29(8): 1023-1026)
- Guo Y.L., Tao B., and Gao X.W., 2008, Characteristic of maize glutathione S-transferase (GSTs) and induction action of acetochlor, *Yimi Kexue (Journal of Maize Sciences)*, 16(1): 122-125 (郭玉莲, 陶波, 高希武, 2008, 玉米谷胱甘肽转

- 移酶(GSTs)特性及除草剂的诱导作用, 玉米科学, 16(1): 122-125)
- George S., Venkataraman G., and Parida A., 2010, A chloroplastlocalized and auxin-induced glutathione S-transferase from phreatophyte *Prosopis juliflora* confer drought tolerance on tobacco, *J. Plant Physiol.*, 167(4): 311-318
- Gao K.H., Ge Y., and Zhang C.H., 2011, Effects of sulfur starvation on the non-protein thiol content and glutathione S-transferase activity of rice seedlings under cadmium stress, *Yingyong Shengtai Xuebao* (Chinese Journal of Applied Ecology), 22(7): 1796-1802 (高可辉, 葛滢, 张春华, 2011, 缺硫对镉胁迫下水稻幼苗非蛋白巯基物质含量和谷胱甘肽硫转移酶活性的影响, 应用生态学报, 22(7): 1796-1802)
- Heikoff S., Greene E.A., Pietrovski S., Bork P., Attwood T.K., and Hood L., 1997, Gene families: the taxonomy of protein paralogs and chimeras, *Science*, 278(5338): 609-614
- Huala E., Dickerman A.W., Garcia-Hernandez M., Weems D., Reiser L., LaFond F., Hanley D., Kiphart D., Zhuang M., Huang W., Mueller L.A., Bhattacharyya D., Bhaya D., Sobral B.W., Beavis W., Meinke D.W., Town C.D., Somerville C., and Rhee S.Y., 2001, The Arabidopsis information resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant, *Nucleic Acids Res.*, 29(1): 102-105
- Huang C., Guo T., Zheng S.C., Feng Q.L., Liang J.H., and Li L., 2009, Increased cold tolerance in *Arabidopsis thaliana* transformed with *Choristoneura fumiferana* glutathione S-transferase gene, *Biol. Plantarum*, 53(1): 183-187
- Itzhaki H., Maxson J.M., and Woodson W.R., 1994, An ethylenesensitive enhancer element is involved in the senescence-related expression of the carnation glutathione-S-transferase (GSTI) gene, *Proc. Natl. Acad. Sci. USA*, 91(19): 8925-8929
- Jain M., Ghanashyam C., and Bhattacharjee A., 2010, Comprehensive expression analysis suggests overlapping and specific roles of rice glutathione S-transferase genes during development and stress responses, *BMC Genomics*, 11: 73-89
- Jain M., Nijhawan A., Arora R., Agarwal P., Ray S., Sharma P., Kapoor S., Tyagi A.K., and Khurana J.P., 2007, F-box proteins in rice, Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress, *Plant Physiol.*, 143(4): 1467-1483
- Ji W., Zhu Y.M., Li Y., Yang L., Zhao X., Cai H., and Bai X., 2010, Over-expression of a glutathione S-transferase gene, *GsGST*, from wild soybean (*Glycine soja*) enhances drought and salt tolerance in transgenic tobacco, *Biotechnol. Lett.*, 32(8): 1173-1177
- Karavangeli M., Labrou N.E., Clonis Y.D., and Tsaftaris A., 2005, Development of transgenic tobacco plants overexpressing maize glutathione S-transferase I for chloroacetanilide herbicides phytoremediation, *Biomol. Eng.*, 22(4): 121-128
- Letunic I., Doerks T., and Bork P., 2012, SMART 7: recent updates to the protein domain annotation resource, *Nucleic Acids Res.*, 40: 302-305
- Li L., Xu H.L., Yang X.L., Li Y.X., and Hu Y.K., 2011, Genome-wide identification, classification and expression analysis of LEA gene family in soybean, *Zhongguo Nongye Kexue* (*Scientia Agricultura Sinica*), 44(19): 3945-3954 (李乐, 许红亮, 杨兴露, 李雅轩, 胡英考, 2011, 大豆 LEA 基因家族全基因组鉴定、分类和表达, 中国农业科学, 44(19): 3945-3954)
- Liu X., He B.W., and Zhang Z.Y., 2009, Stress-responsive expression analysis of glutathione-S-transferase gene of *Ipomoea batatas* (L.) Lam, *Yichuan* (*Hereditas* (Beijing)), 31(8): 859-864 (刘珣, 何博文, 张义正, 2009, 甘薯谷胱甘肽-S-转移酶基因在胁迫条件下的表达分析, 遗传, 31(8): 859-864)
- Long H.T., Li L., and Wan X.R., 2004, Relationship between ABA-inducible gene and stress, *Yaredai Zhiwu Kexue* (*Subtropical Plant Science*), 33(4): 74-77 (龙海涛, 李玲, 王小荣, 2004, ABA 诱导基因及其与逆境胁迫的关系, 亚热带植物科学, 33(4): 74-77)
- McGonigle B., Keeler S.J., Lau S.M., Koeppe M.K., and O'Keefe D.P., 2000, A genomics approach to the comprehensive analysis of the glutathione S-transferase gene family in soybean and maize, *Plant Physiol.*, 124(3): 1105-1120
- Moons A., 2003, *Osgstu3* and *Osgtu4*, encoding tau class glutathione S-transferases, are heavy metal and hypoxic stress-induced and differentially salt stress-responsive in rice roots, *FEBS Lett.*, 553(3): 427-432
- Nakashima K., Fujita Y., Katsura K., Maruyama K., Narusaka Y., Seki M., Shinozaki K., and Yamaguchi-Shinozaki K., 2006, Transcriptional regulation of ABI3- and ABA-responsive genes including *RD29B* and *RD29A* in seeds, germinating embryos, and seedlings of *Arabidopsis*, *Plant Mol. Biol.*, 60(1): 51-68
- Pascal S., Gullner G., Kömives T., and Scalla R., 2000, Selective induction of glutathione S-transferase subunits in wheat plants exposed to the herbicide acifluorfen, *Z. Naturforsch. C.*, 55: 37-39
- Qin F., Sakuma Y., Li J., Liu Q., Li Y.Q., Shinozaki K., and Yamaguchi-Shinozaki K., 2004, Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L., *Plant Cell Physiol.*, 45(8): 1042-1052
- Schmutz J., Cannon S.B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D.L., Song Q., Thelen J.J., Cheng J., Xu D., Hellsten U., May G.D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M.K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X.C., Shinozaki K., Nguyen H.T., Wing R.A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R.C., and Jackson S.A., 2010, Genome sequence of the palaeopolyploid soybean, *Nature*, 463(7278): 178-183
- Sheehan D., Meade G., Foley V.M., and Dowd C.A., 2001, Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily, *Biochem J.*, 360(1): 1-16
- Shinozaki K., and Yamaguchi-Shinozaki K., 2000, Molecular responses to dehydration and low temperature: differences and cross-talk between two stress signaling pathways, *Curr. Opin. Plant Biol.*, 3(3): 217-223
- Tamura K., Peterson D., Peterson N., Stecher G., Nei M., and Kumar S., 2011, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol.*, 28(10): 2731-2739

- Thom R., Dixon D.P., Edwards R., Cole D.J., and Laphorn A.J., 2001, The structure of a zeta class glutathione S-transferase from *Arabidopsis thaliana*: characterization of a GST with novel active site architecture and a putative role in tyrosine catabolism, *J. Mol. Biol.*, 308(5): 949-962
- Wagner U., Edwards R., Dixon D.P., and Mauch F., 2002, Probing the diversity of the *Arabidopsis* glutathione S-Transferase gene family, *Plant Mol. Biol.*, 49(5): 515-532
- Wang Z.Y., Cai H., Bai X., Ji W., Li Y., Wei Z.W., and Zhu Y.M., 2012, Isolation of GsGST19 from *Glycine soja* and analysis of saline-alkaline tolerance for transgenic *Medicago sativa*, *Zuowu Xuebao (Acta Agronomica Sinica)*, 38(6): 971- 979 (王臻昱, 才华, 柏锡, 纪巍, 李勇, 魏正巍, 朱延明, 2012, 野生大豆 GsGST19 基因的克隆及其转基因苜蓿的耐盐碱性分析, *作物学报*, 38(6): 971-979)
- Wilce M.C., and Parker M.W., 1994, Structure and function of glutathione S-transferases, *Biochim. Biophys. Acta*, 1205(1): 1-18
- Xiang X., and Fu J.R., 1998, The expression and regulation of ABA-responsive genes and the relation with stress, *Zhoxue Tongbao (Chinese Bulletin of Botany)*, 15(3): 11-16 (向旭, 傅家瑞, 1998, 脱落酸应答基因的表达调控及其与逆境胁迫的关系, *植物学通报*, 15(3): 11-16)
- Yang S., Zhang X., Yue J.X., Tian D., and Chen J.Q., 2008, Recent duplications dominate NBS-encoding gene expansion in two woody species, *Mol. Genet. Genomic*, 280(3): 187-198
- Zhao F.Y., Wang X.Y., Zhao Y.X., and Zhang H., 2006, Transferring the *Suaeda salsa* glutathione S-transferase and catalase genes enhances low temperature stress resistance in transgenic rice seedlings, *Zhiwu Shengli Yu Fenzi Shengwu Xuebao (Journal of Plant Physiology and Molecular Biology)*, 32(2): 231-238 (赵风云, 王晓云, 赵彦修, 张慧, 2006, 转入盐地碱蓬谷胱甘肽转移酶和过氧化氢酶基因增强水稻幼苗对低温胁迫的抗性, *植物生理与分子生物学学报*, 32(2): 231-238)
- Zheng W.J., Xu Z.S., Feng Z.J., Li L.C., Chen M., Chai S.C., and Ma Y.Z., 2012, Genome-wide identification, classification, and expression of NF-YB gene family in soybean, *Zuowu Xuebao (Acta Agronomica Sinica)*, 38(9): 1570-1582 (郑炜君, 徐兆师, 冯志娟, 李连城, 陈明, 柴守诚, 马有志, 2012, 大豆 NF-YB 家族全基因组鉴定、分类和表达, *作物学报*, 38(9): 1570-1582)
- Zhu Q.S., Huang J.Z., Zhou Y., Ye Q.F., Xu B.J., and Chen Z.Y., Herbicide-induced oxidative stress in seedlings of two rice cultivars with different chilling tolerance, *Henongxue Bao (Acta Agriculturae Nucleatae Sinica)*, 23(1): 145-149 (朱其松, 黄建中, 周焯, 叶庆富, 徐步进, 陈子元, 2009, 除草剂对不同耐寒性水稻幼苗的氧化胁迫效应, *核农学报*, 23(1): 145-149)