

玉米自交系 B73 全基因组 cystatin 抗虫基因家族分析

于金海

(黑龙江省农业科学院 作物育种研究所, 黑龙江 哈尔滨 150086)

摘要:半胱氨酸蛋白酶抑制剂(cystatin)基因具有 cystatin 结构特征的蛋白结构域,广泛存在于植物或动物体内。此类基因在植物或动物蛋白质组中具有多个功能各异的家族成员,并在抗虫中扮演重要角色。通过利用玉米 B73 全基因组的基因表达序列数据,对 12 个 cystatin 基因的结构组成、染色体分布、系统进化等进行了分析。鉴定了该基因家族在蛋白序列间 3 个高度保守基序,构建了基因家族系统进化树,为该重要基因家族的进化提供了依据的同时也为克隆和分离该类抗虫基因提供参考。

关键词:玉米;cystatin 基因;全基因组;抗虫;基因家族

中图分类号:S513

文献标识码:A

文章编号:1002-2767(2010)08-0010-04

虫害严重影响农作物的产量和品质,制约农业经济的发展。长期以来人们普遍采用化学杀虫的方法来控制害虫,但由于农药的大量使用也带来了农药残留、害虫产生抗药性、自然环境被污染等严重问题。为避免这些缺点,育种家尝试利用转基因方法来提高农作物的抗虫性。自从 1983 年首次获得转基因植物以来,植物抗虫基因工程育种已成为作物抗虫育种研究的热点。目前,已经分离克隆的多种抗虫基因,主要为蛋白酶抑制剂基因^[1-5]。

蛋白酶抑制剂是一类具有抑制蛋白酶活性的物质,植物蛋白酶抑制剂是一种能够抑制蛋白水解酶活性的小分子蛋白或多肽。根据作用于酶的活性基团不同又可将植物中的蛋白酶抑制剂基因分为 4 类,即丝氨酸类、半胱氨酸类、天门冬氨酸类和金属类。其中,半胱氨酸类蛋白酶抑制剂(cystatin)对以半胱氨酸类蛋白酶为主要消化酶的鞘翅目昆虫的防治具有重要作用^[6]。而由于半胱氨酸蛋白酶抑制剂不影响哺乳动物包括人类消化酶的活性,因而备受青睐。目前,玉米和水稻的 cystatin 基因的 cDNA 都被克隆^[7-10],而且二者具有很高的同源性。

通过植物来源的抗虫基因间与其它来源的抗虫基因的联合使用,培育抗虫作物提高抗虫性已成为现实^[11-13]。不断发现和克隆新的抗虫基因将是今后植物抗虫育种领域研究的重点,由于农作

物本身天然抗虫机制的存在,作物本身将会是今后人们发掘更多抗虫基因的重要来源。尤其是随着全基因组测序计划的完成^[14],为发掘更多的抗虫基因提供了可能。该研究采用生物信息学方法,利用玉米自交系 B73 全基因组测序结果,全面分析了半胱氨酸蛋白酶抑制剂(cystatin)基因家族的分布、结构以及系统进化等,为研究该基因家族的进化和分离新功能基因提供参考。

1 材料与方法

1.1 基因表达数据的获得

以玉米自交系 B73 全基因组表达序列为研究对象,全基因组数据库来自玉米测序网站 <http://www.maizesequence.org/index.html>,共获得了玉米 53 764 个基因的表达序列、拟翻译蛋白序列以及基因全长序列。利用 Interproscan 软件预测这些基因所编码的蛋白序列的结构域,共获得玉米半胱氨酸蛋白酶抑制剂基因 12 个,同时提取基因的全长序列和蛋白序列。

1.2 序列分析

1.2.1 蛋白结构域分析 利用 MEME/MAST 程序(Motif discover and search V4.3.0, http://meme.sdsc.edu/meme4_3_0/cgi-bin/meme.cgi)^[15],对玉米 Cystatin 蛋白全序列进行分析,鉴定蛋白序列包含的基序(motif),基序数目限制在 30 个,基序长度限制在 20~100 个氨基酸残基,单个结构域在序列间的分布取任意值。

1.2.2 构建系统进化树 采用 CLUSTAL W(V1.83)程序对氨基酸全序列进行联配,参数默认。采用邻位相连法构建系统进化树,并利用 TreeView 软件显示系统发生树。

收稿日期:2010-04-26

作者简介:于金海(1971-),男,黑龙江省密山市人,学士,助理研究员,从事玉米遗传育种研究。E-mail: hnygy@163.com。

2 结果与分析

2.1 玉米 cystatin 基因的鉴定

从玉米 B73 中鉴定的 12 个 cystatin 基因分别命名为 *ZmCYS1*~12,12 个 *ZmCYS* 基因在玉米全基因组中的分布情况及基因长度、外显子数目见表 1,可知,玉米 *ZmCYS* 基因长度介于 554~6211 bp,基因长度最大的是 *ZmCYS4* 基因,为 6 211 bp,长度最小的是 *ZmCYS12*,为 554 bp。从基因在染色体分布看,*ZmCYS* 基因家族成员在基因组中分布并不均匀,其中,玉米 *ZmCYS* 基因在第 8 染色体分布最多,有 4 个基因位于该染色体上;其次是第 1 染色体,有 3 个基因成簇分布在该染色体上。第 6

染色体上有 2 个 *ZmCYS* 基因,第 3、4、9 染色体上各有 1 个 *ZmCYS* 基因。从基因结构看,8 个(66.7%)*ZmCYS* 基因的外显子数为 1 个或 2 个,外显子数最多的为 *ZmCYS9* 基因,为 4 个,其余 3 个基因有 3 个外显子。4 个 *ZmCYS* 基因还存在可变剪切情况(见表 2),其中 *ZmCYS4*、*ZmCYS6* 和 *ZmCYS11* 存在 2 种可变剪切,*ZmCYS4* 和 *ZmCYS11* 基因可变剪切产物相同,长度分别为 135 和 176 个氨基酸,而 *ZmCYS6* 基因 2 种可变剪切产生的蛋白产物不同,分别为 225 和 245 个氨基酸。*ZmCYS9* 基因存在 4 种可变剪切,产生 3 种蛋白产物,分别为 150、164 和 173 个氨基酸。

表 1 玉米 B73 基因组中 cystatin 基因及其基本信息

基因名称	染色体	基因 ID	转录起始	转录终止	基因长度/bp	外显子数目/个
<i>ZmCYS1</i>	Chr1	GRMZM2G148925	77866901	77868393	1492	1
<i>ZmCYS2</i>	Chr1	GRMZM2G401328	77931069	77931873	804	1
<i>ZmCYS3</i>	Chr1	GRMZM2G401374	77938590	77939262	672	1
<i>ZmCYS4</i>	Chr3	GRMZM2G438551	185990263	185996474	6211	2
<i>ZmCYS5</i>	Chr4	GRMZM2G312061	15234221	15235083	862	2
<i>ZmCYS6</i>	Chr6	GRMZM2G024264	146643037	146645456	2419	3
<i>ZmCYS7</i>	Chr6	GRMZM2G440968	155818803	155820790	1987	2
<i>ZmCYS8</i>	Chr8	GRMZM2G012160	170150380	170153072	2692	2
<i>ZmCYS9</i>	Chr8	GRMZM2G013461	3326253	3329125	2872	4
<i>ZmCYS10</i>	Chr8	GRMZM2G030717	161282967	161284061	1094	3
<i>ZmCYS11</i>	Chr8	GRMZM2G133620	104172564	104177448	4884	3
<i>ZmCYS12</i>	Chr9	GRMZM2G065259	146736785	146737339	554	1

表 2 cystatin 基因存在的可变剪切

基因名称	剪切位点序号	转录起始	转录终止	氨基酸长度
<i>ZmCYS4</i>	1	185990263	185996474	135
	2	185990263	185996041	135
<i>ZmCYS6</i>	1	3326253	3329125	245
	2	3326382	3329125	225
<i>ZmCYS9</i>	1	146643037	146645456	164
	2	146643050	146645456	150
	3	146643050	146644346	173
	4	146643050	146644269	164
<i>ZmCYS11</i>	1	104172564	104177448	176
	2	104173753	104177096	176

2.2 玉米 cystatin 蛋白的结构域

从一级结构上看,12 个 *ZmCYS* 基因编码的蛋白质差别不大,具有 105~245 个氨基酸残基。对这些氨基酸进行基序分析表明,这 12 个基因所编码的蛋白主要包含 3 种基序,这些基序的结构见图 1,这 3 个基序提取到的一致序列见表 3,其中基序 1 中含有 cystatin 基因标志性基序“QVVAG”,这也是该基因家族的活性位点。从图 1 可以看出,*ZmCYS*1 等 8 个基因包含全部 3 种基序,排列顺序为 3-1-2。*ZmCYS*6、*ZmCYS*9 和 *ZmCYS*11 基因只含有 1、2 基序,缺失基序 3。而 *ZmCYS*5 则含有基序 3 和 1,缺失基序 2。这 3 个基序所在位置基本对应 cystatin 的基本结构域,*ZmCYS*9 基因含有 2 个 cystatin 基本结构域,其它 11 个基因都只含有 1 个。9 个基因的蛋白起始区包含 23~33 个氨基酸残基的信号肽,而 *ZmCYS*2、*ZmCYS*7 和 *ZmCYS*10 则不包含信号肽,*ZmCYS*10 基因在 58~80 处氨基酸残基处有一个跨膜区域。

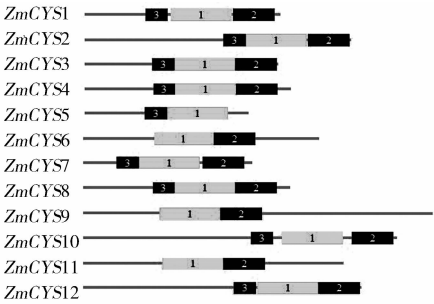


图 1 Cystatin 蛋白保守基序分布图

表 3 MEME/MAST 分析提取的 cystatin 蛋白保守基序

基序编号	共有序列
1	RFAVAEHNKKANAGLEFERVVKGKEQV-VAGTLYDLTLEAKD
2	AGKKKLYEAKVWEKPWEDFKELQEFAPV
3	NENDPHIQELG

2.3 玉米 cystatin 蛋白的系统发生分析

为了解玉米 cystatin 基因的系统进化关系,利用 Clustal W 对该基因家族的 12 个成员所编码的 18 种蛋白进行了系统发生分析,构建了系统发生树(见图 2)。根据图 2 所呈现的进化关系的远近,可将玉米 cystatin 基因分成 4 个亚族,分别用 G1~G4 表示。其中,第 1 亚族(G1)中包含的成员最多,共含有 4 个基因分别为 *ZmCYS*4、*ZmCYS*5、*ZmCYS*8 和 *ZmCYS*11 的 6 个蛋白序列。第 2 亚族包含 *ZmCYS*1、*ZmCYS*3 和 *ZmCYS*7。第 3 亚族包含 *ZmCYS*6 和 *ZmCYS*9 基因的 6 个蛋白,可见同一基因不同剪切所得蛋白亲缘关系较近。第 4 亚族包含 3 个成员,分别为 *ZmCYS*2、*ZmCYS*10 和 *ZmCYS*12。

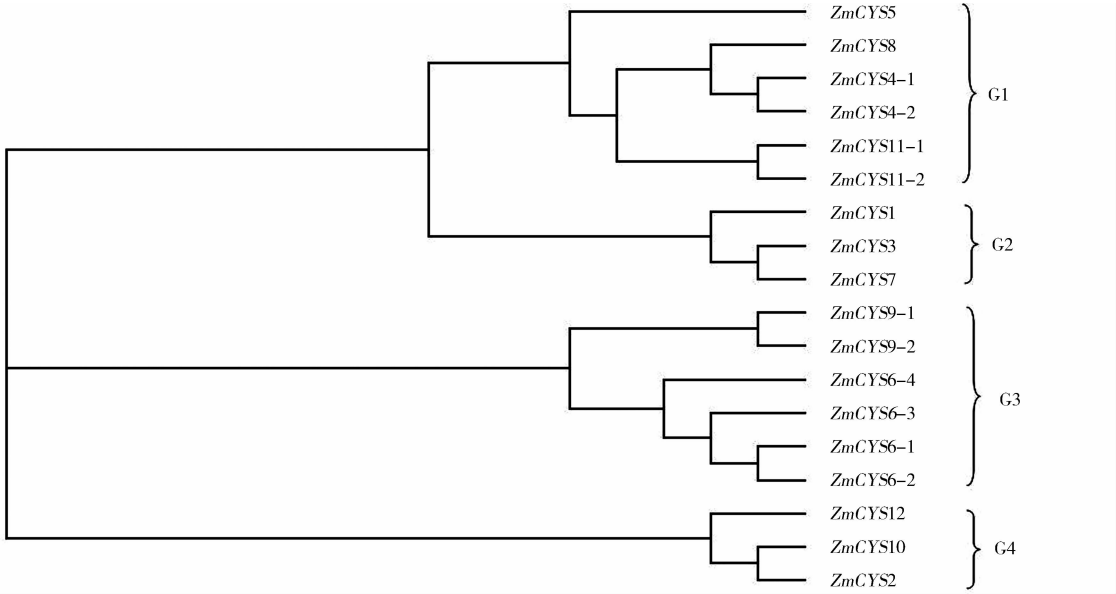


图 2 玉米 Cystatin 蛋白系统发生树

3 结论与讨论

利用生物信息学方法,在玉米自交系 B73 的全基因组中共找到 12 个 cystatin 基因,这与在水稻全基因组中发现的 cystatin 基因的数量一致^[16],数量上明显少于抗病基因^[17]。根据玉米 cystatin 基因编码蛋白的系统发生树分析可以看出,位于同一亚族的基因序列相互间的距离均小于它们与其它序列之间的距离,被认为具有相似结构与功能的直系同源簇^[18]。玉米 cystatin 基因家族在长期进化过程中已表现出结构上的一些差异,个别基因内部也发生了复制现象。

该研究利用多重联配以及 MEME 分析表明,cystatin 基因蛋白家族中存在高度保守的基序,尤其是 QVVxG 基序在所有的 cystatin 基因中非常保守。这些不同的保守基序可能与基因的特定功能相关。

此外,对 cystatin 基因的表达情况进行了探索,在 GenBank 数据库中发现 26 个 cystatin 基因的表达序列标签,分别来自种子、胚、叶片和根,这些组织是受昆虫侵害较多的部位,在这些部位表达有助于增强对昆虫的抗性。该研究对所鉴定的 cystatin 基因家族结构进行的系统分析将有助于对该基因的克隆和分离,对玉米抗虫育种具有参考价值。

参考文献:

- [1] Vaeck M, Reynaerts A, Herman H, et al. Transgenic plants protected from insect attack. [J]. Nature, 1987, 328: 33-37.
- [2] Hilder V A, Gatehouse A M R, Sheerman S E, et al. A novel mechanism of insect resistance engineered into tobacco [J]. Nature, 1987, 33: 160-163.
- [3] 高越峰, 朱祯, 肖桂芳, 等. 大豆 Kunitz 型胰蛋白酶抑制剂基因的分离及其在抗虫植物基因工程中的应用 [J]. 植物学报, 1998, 40(5): 404-406.
- [4] Rao K V, Rathore K S, Hodges T K, et al. Expression of snowdrop lectin (GNA) in transgenic rice plants confers resistance to rice brown planthopper [J]. Plant Journal, 1998, 15(4): 469-477.
- [5] 梁辉, 朱银峰, 朱祯, 等. 雪莲花凝集素基因转化小麦及转基因小麦抗蚜性的研究 [J]. 遗传学报, 2004, 31(2): 189-195.
- [6] Ryan C A. Proteinase inhibitor gene families, strategies for transformation improve plant defense against herbivores [J]. Bioessays, 1989, 10(1): 20-22.
- [7] Chen M S, Johnson B, Wen L, et al. Rice cystatin: bacterial expression, purification, cysteine proteinase inhibitory activity, and insect growth suppressing activity of a truncated form of the protein [J]. Protein Expr Purif, 1992, 3(1): 41-49.
- [8] Takafumi Y, Hiroyuki O, Azusa S, et al. A Cysteine Protease from Maize Isolated in a Complex with Cystatin [J]. Plant and Cell Physiology, 2000, 41(2): 185-191.
- [9] Massonneau A, Condamine P, Wisniewski J P, et al. Maize cystatins respond to developmental cues, cold stress and drought [J]. Biochim Biophys Acta, 2005, 1729(3): 186-199.
- [10] Sadami O, Mayumi T, Tomokazu K, et al. Oryza cystatin-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 Oryza cystatin-II [J]. Journal Agric Food Chem, 2007, 55(5): 1762-1766.
- [11] 卢晓风, 夏玉先, 裴炎. 植物蛋白酶抑制剂在植物抗虫与抗病中的作用 [J]. 生物化学与生物物理学进展, 1998, 25(4): 328-333.
- [12] 柳武革, 薛庆中. 蛋白酶抑制剂及其在抗虫基因工程中的应用 [J]. 生物技术通报, 2000(1): 20-25.
- [13] 达克东, 崔德才, 张松, 等. 超强表达豇豆胰蛋白酶抑制剂基因 (*CpTI*) 转化苹果的研究 [J]. 园艺学报, 2001, 28(1): 57-58.
- [14] Patrick S, Schnable D W, Robert S, et al. The B73 Maize Genome: Complexity, Diversity, and Dynamics [J]. Science, 2009, 326: 1112-1115.
- [15] Bailey T L, Gribskov M. Combining evidence using P-values; application to sequences homology searches [J]. Bioinformatics, 1998, 14: 48-54.
- [16] 杨泽峰, 顾世梁, 许花, 等. 拟南芥和水稻 cystatin 基因家族的生物信息学分析 [J]. 扬州大学学报, 2007, 28(3): 51-57.
- [17] 汪结明, 江海洋, 赵阳, 等. 玉米自交系 B73 全基因组 NBS 类型抗病基因分析 [J]. 作物学报, 2009, 35(3): 566-570.
- [18] Ledent V, Paquet O, Venvoort M. Phylogenetic analysis of the human basic helix-loop-helix proteins [J]. Genome Biology, 2002, 3(6): 30-38.

Genome-Wide Analysis of cystatin Gene Family in Maize Inbred Line B73

YU Jin-hai

(Crop Breeding Institute of Heilongjiang Academy of Agricultural Sciences, Harbin, Heilongjiang 150086)

Abstract: Cystatin genes are cysteine proteinase inhibitors with a protein pattern of cystatin-containing domain, which exist widely in many proteins from plants and animals. This protein has many different family members distributing in plant and animal proteome, and plays key roles in plants against pests. Here, we identified complete set of 12 different cystatin gene in the maize (*Zea mays* L.) inbred line B73 genome. The putative cystatin genes were characterized with respect to structural diversity, chromosome distribution, phylogenetic relationship and so on. Three conserved motifs were found among proteins of cystatin gene of maize. Phylogenetic analysis of these proteins indicated that four groups exist. The results of this study will provide evidence for evolution for cystatin gene family in maize and benefit for separating and cloning resistance gene to defend the insects.

Key words: maize; cystatin gene; genome wide; insect-resistance; gene family