

胡军华,王大伟,杨明磊,等. 番茄 C2H2 型锌指蛋白的鉴定及生物信息学分析[J]. 江苏农业科学,2018,46(12):23-27.

doi:10.15889/j.issn.1002-1302.2018.12.005

番茄 C2H2 型锌指蛋白的鉴定及生物信息学分析

胡军华^{1,2},王大伟¹,杨明磊^{1,2},王新^{1,2},朱海²,朱佩²,刘贯山¹

(1. 中国农业科学院烟草研究所/烟草行业烟草基因资源利用重点实验室,山东青岛 266101; 2. 中国农业科学院研究生院,北京 100081)

摘要:植物 C2H2 型锌指蛋白转录因子广泛参与植物的生长发育及对逆境应答反应等生命过程。为了更好地理解番茄中 C2H2 型锌指蛋白转录因子的种类和数量,利用番茄基因组数据库,鉴定出 92 条番茄 C2H2 型锌指蛋白。蛋白质理化分析结果表明,不同亚家族成员的氨基酸长度差异较大,且多为碱性氨基酸;进化树分析结果表明,所有 C2H2 家族成员被分为 5 个亚家族,每个亚家族成员数量差异较大;结构域分析显示每个成员都含有 C2H2 结构域,同一亚家族间结构域的数量、种类、分布具有一致性;染色体定位结果显示,有 91 个成员定位在 12 条染色体上;组织表达分析结果表明,C2H2 转录因子基因家族成员在各个组织中都有表达,表达量具有差异性。本研究将为番茄 C2H2 型锌指蛋白的生物学功能分析提供参考。

关键词:C2H2 型锌指蛋白;番茄;生物信息学;进化树;结构域;染色体定位;组织表达

中图分类号:S641.201 **文献标志码:**A **文章编号:**1002-1302(2018)12-0023-05

C2H2 型锌指蛋白是最常见的锌指蛋白之一,在真核生物中发挥重要的生物学功能^[1]。根据已有报道,对拟南芥及杨树等植物中的 C2H2 型锌指蛋白序列分析表明,多数成员含有保守氨基酸序列 C-X2~4-C-X3-P-X5-L-X2-H-X3-H(其中 C 表示半胱氨酸,H 表示组氨酸,X 表示任意氨基酸,P 表示苯丙氨酸,L 表示亮氨酸)^[2]。在锌指结构中还存在 1 个非常保守的 QALGGH 序列,它在转录因子对下游 DNA 序列的结合中发挥重要作用^[3]。这些氨基酸序列与 Zn²⁺ 一起形成紧密的 $\beta\beta\alpha$ 四面体结构^[4]。一些研究结果表明,只有单锌指结构的 C2H2 型锌指蛋白转录因子可能是与植物生长发育,尤其是生殖器官发育有关。例如,拟南芥中的 SUPERMAN 基因突变,会导致拟南芥雄蕊过度分裂^[5]。而有多个 C2H2 型锌指蛋白的转录因子则更多地参与各种植物胁迫,如拟南芥中的双锌指蛋白 AZF1、AZF2、AZF3,能提高植物对盐的胁迫能力^[6]。

C2H2 型锌指蛋白在其他植物中的生物信息学分析已有发表,杨明磊等从烟草中鉴定出了 116 条 C2H2 型锌指蛋白家族成员^[7]。番茄全基因组测序工作在 2012 年就已经完成并开放,此后,迅速地推动了对番茄基因组的生物信息学的分析研究^[8]。李晓旭等运用番茄全基因组信息完成 WOX 转录因子家族的鉴定及其进化、表达分析^[9]。阮美颖等借助番茄全基因组测序数据鉴定了查尔酮合成酶基因家族成员^[10]。番茄作为茄科植物功能基因组学研究的重要模式生物,目前

还没有其 C2H2 型锌指蛋白的信息学分析。本研究以番茄 C2H2 型锌指蛋白为研究对象,采用生物信息学手段对其物理化学性质、进化关系、保守结构域、三维结构、表达情况等方面进行分析,为番茄的 C2H2 型锌指蛋白的深入研究提供了基础。

1 材料与方法

1.1 蛋白序列的鉴定与获得

根据有关 C2H2 蛋白的文献,在 TAIR 数据库中,下载拟南芥的 C2H2 型锌指蛋白全长序列,再从 Pfam 数据库中下载 C2H2 型锌指蛋白种子序列(Pfam ID:PF00096)。然后将 2 组序列在番茄数据库中进行 Blastp 比对(E 值 $\leq 1 \times 10^{-15}$)。所得结果在 Pfam 数据库中对 C2H2 结构域进行预测,淘汰不含 C2H2 的序列。利用 ExPASy Proteomics Server 在线工具预测蛋白质的基本信息,包括分子量、等电点等。利用 WoLF PSORT 在线软件的对亚细胞定位分析^[11]。

1.2 进化树的构建

利用 MUSCLE 软件将番茄 C2H2 家族成员的蛋白全长序列进行多序列比对,比对结果通过 MEGA7 软件构建邻接树,检验次数 Bootstrap 重复 1 000 次^[12]。

1.3 三维结构域的预测与分析

利用 MEM 在线工具进行蛋白序列 motif 搜索,下载保守序列并将结果可视化^[13]。在 SWISS-MODEL 上查找三维结构模板,选择最适三维结构模板,使用 Pymol 软件标出保守结构域。

1.4 染色体定位

利用 Perl 程序解析番茄基因组信息文件,确定 C2H2 基因家族成员在染色体上的位置,通过 MapGene2 软件绘制定位图谱^[14],根据图谱绘制统计表格。

1.5 基因家族的表达分析

在 NCBI 中下载番茄转录组测序数据,通过 R 语言 Bioconductor 程序对转录组数据进行分析,鉴定并提取 C2H2

收稿日期:2016-12-07

基金项目:国家烟草专卖局烟草基因组计划重大专项[编号:110201301005(JY-05)];中国农业科学院烟草研究所青年科学基金(编号:2015B05)。

作者简介:胡军华(1990—),男,安徽黄山人,硕士研究生,主要从事烟草突变体鉴定与利用方向的研究。E-mail:1577209402@qq.com。

通信作者:刘贯山,研究员,博士生导师,主要从事烟草分子育种研究。E-mail:liuguanshan@caas.cn。

家族转录组数据,使用 R 语言 gplots 程序包绘制热图。

2 结果与分析

2.1 番茄 C2H2 基因家族成员的鉴定及蛋白质理化性质、亚细胞定位分析

通过筛选鉴定,最终获得 92 条番茄 C2H2 锌指蛋白序列。将它们进行蛋白质理化性质分析,结果(表 1)显示,C2H2 基因家族蛋白质长度介于 96~673 个氨基酸之间,蛋白

质分子量介于 11 006.4~75 310.4 u,同一亚家族中的氨基酸长度也有较大差异。番茄 C2H2 基因家族所有成员的蛋白质序列等电点 PI 介于 5.52~10.01 之间,变化较大,但是同一亚家族的蛋白质成员等电点变化较小,亚家族 IV 蛋白质成员等电点多集中 8.31~9.38 之间,亚家族 III 蛋白质成员等电点均大于 9.5,总体上看,番茄 C2H2 型锌指蛋白多数为碱性氨基酸。亚细胞定位分析表明,95% 以上的家族成员在细胞核中起作用,直接证明了该基因家族所鉴定蛋白序列的可靠性。

表 1 番茄 C2H2 锌指蛋白家族理化特征分析

亚家族	个数	分子量 (u)	氨基酸数目 (个)	理论等电点		亚细胞定位			
				≥7	<7	nucl	cysk	cyto	plas
I	30	11 006.4~40 888.7	96~360	20	10	29	1	0	0
II	25	17 369.1~75 310.4	152~673	14	11	24	0	0	1
III	3	15 593.8~17 092.4	137~152	3	0	3	0	0	0
IV	5	17 444.8~39 386.3	153~364	5	0	5	0	0	0
V	29	24 611.7~68 007.6	211~656	23	6	28	0	1	0

2.2 番茄 C2H2 型锌指蛋白的进化分析及结构域分布

为揭示番茄 C2H2 型锌指蛋白成员之间的进化关系,将 92 条番茄 C2H2 型锌指蛋白全长序列进行多序列比对,根据系统进化树将番茄 C2H2 型锌指蛋白成员分为 5 个亚家族(I、II、III、IV、V),如图 1 所示,每个亚家族分别有 30、25、3、5、29 个成员。进化树分析显示,进化树支持率高于 96% 的姊妹对有 30 对。利用 MEME 对番茄 C2H2 型锌指蛋白进行结构域分析,并绘制图谱。结果显示所有该基因家族成员都含有 C2H2 锌指结构域(Motif1),同一亚家族每个蛋白质序列含有的结构域数量、种类及其分布位置具有较高的一致性。亚家族 I、III、IV、V 绝大部分蛋白质序列仅含有 1 个 C2H2 锌指结构域,且多靠近于肽链的 N-端。亚家族 II 大多数成员蛋白质序列仅含有 Motif1、Motif2,所含结构域数量上也较一致,多数蛋白序列含有 2 个 C2H2 锌指结构域。亚家族 V 的所有成员所含有的结构域的数量及种类都较其他亚家族多,该亚家族成员可以分为 2 类:一类不含有 Motif5,只有 1 条蛋白质序列,只含有 1 个结构域 Motif1,其他大部分蛋白序列含有 3 种以上的结构域;另一类包含 Motif5 的蛋白序列,大部分包含 5 个以上不同的结构域。以上分析表明,每个亚家族所包含的结构域在数量和种类上有较高的一致性,从侧面反映了进化树的可靠性。

2.3 番茄 C2H2 锌指蛋白结构域特征分析

利用 MEME 软件对全长蛋白质序列预测,得到 6 个结构域(图 2),对每个结构域进行分析发现,Motif1 中含有 29 个比较保守的氨基酸序列,且存在 C2H2 结构域。其他 Motif 不存在 C2H2 结构域。Motif2 中含有 21 个保守氨基酸,其中有 C-X2-C-X12-H 结构序列,X 保守性不高。Motif3 有 41 个保守氨基酸,且保守性较高。Motif4、Motif5、Motif6 保守氨基酸数目较少,但保守性较高。

通过多序列比对发现,所有番茄上都含有 C-X2~4-C-X3-P-X5-L-X2-H-X3-H,即 C2H2 结构域,且也都存在 QALGGH 序列,这与其他作物研究结果一致。从三维空间结构图(图 2)可以看出,中心位置 Zn²⁺ 与 2 个 His 和 2 个 Cys 相结合形成紧密的 $\beta\beta\alpha$ 四面体结构以维持整个锌指结构的稳定。

2.4 番茄 C2H2 基因家族染色体定位分析

利用番茄基因组数据库,将 92 条番茄 C2H2 转录因子成员定位在 12 条染色体上,定位结果显示,92 个 C2H2 基因中有 91 个基因分布于 12 条染色体中,且呈不均匀分布(表 2)。6 号染色体上分布最多,共有 13 个家族成员;1 号和 5 号染色体分布次之,分别含有 12 个和 10 个家族成员;12 号染色体上分布最少,仅含有 1 个候选基因。亚家族 I 有 1 个基因(Solyc00g015730.1.1)无法确定在染色体上的位置,只定位在 1 个 scaffold 上,即 0 号染色体上。

通过各个亚家族在染色体上的分布定位图还发现,同一家族各成员在染色体上的分布也存在差异。亚家族 V 的各个成员分布比较均匀,共有 29 个基因定位在染色体上,且在 11 条染色体上均有分布,只在 12 号染色体无分布;亚家族 I 共有 29 个候选基因定位在 10 条染色体上,只在 2 号、12 号没有分布。亚家族 II 的 25 个基因定位在 10 条染色体上,只在 2 号、10 号染色体上无分布。家族 III、IV 本身基因数量较少,分布也较集中,都主要在 2 号染色体上。

2.5 番茄 C2H2 基因家族的组织表达分析

在 NCBI 中下载番茄转录组测序数据,通过 R 语言进行分析鉴定,使用 R 语言 gplots 程序包绘制热图,一共筛选到 11 个 C2H2 基因家族成员的表达数据。将数据进行可视化处理后可以看出:I 家族的 Solyc01g107430.2 在各个组织都有较高表达量,Solyc10g049870 只在根部和种子中表达量较高;II 家族有 2 个成员在各组织表达量较高,另 2 个在各组织表达量则较低;V 家族的 4 个成员整体表达量比较均匀,与其他家族不同的是,它在根部的表达量较低(图 3)。

3 讨论

C2H2 型锌指蛋白是最常见的锌指蛋白之一,对真核生物的生长发育及逆境胁迫的耐受能力都有着重要作用。近些年来,国内外有关植物 C2H2 型锌指蛋白的研究逐步增多,并发现了多种与植物生长发育及胁迫耐受性相关的 C2H2 型锌指蛋白^[15]。

在本研究中,鉴定了 92 条 C2H2 型锌指蛋白成员,95% 以上的家族成员定位在细胞核中,直接证明了该基因家族所

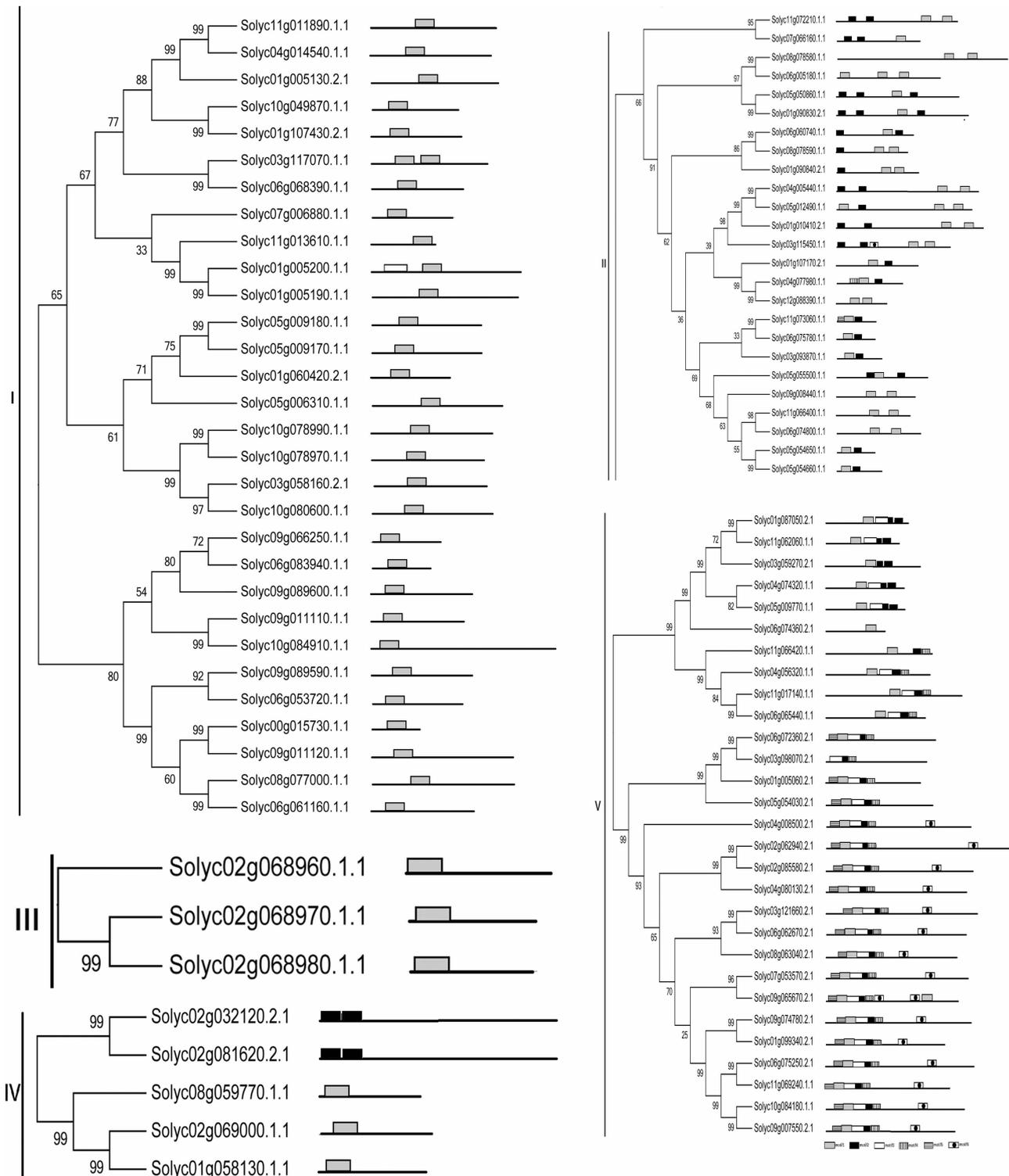


图1 番茄 C2H2 锌指蛋白进化树及结构域分布

鉴定蛋白序列的可靠性。92 个成员能分成 5 个亚家族, 各亚家族间数量差异较大, 同一亚家族中的各个成员所含有结构域的数目和种类高度一致, 蛋白质序列相似度较高, 不同亚家族之间差异明显, 表明不同亚家族成员可能具有不同的生物学功能, 这也验证了进化树的可靠性。

植物中不同的锌指蛋白转录因子中所含有的锌指蛋白数目不同, 一般 1 个到 4 个不等, 含有不同个数锌指蛋白转录因

子在植物中有各自不同的作用机制。番茄结构域分析结果表明, 大多数成员只含有 1 个 C2H2 型锌指蛋白, 亚家族 II 含有 2 个 C2H2 型锌指蛋白。单锌指蛋白具备了对靶 DNA 的转录调节功能, 双锌指蛋白锌指区的距离决定了靶 DNA 结合区域的大小和序列的专一性识别, 与其他亚家族相比, II 亚家族可能更具有特异性作用^[16]。在 C2H2 结构域中存在 QALGGH 基序, 为植物所特有, 对于 C2H2 型锌指蛋白转录因子结合

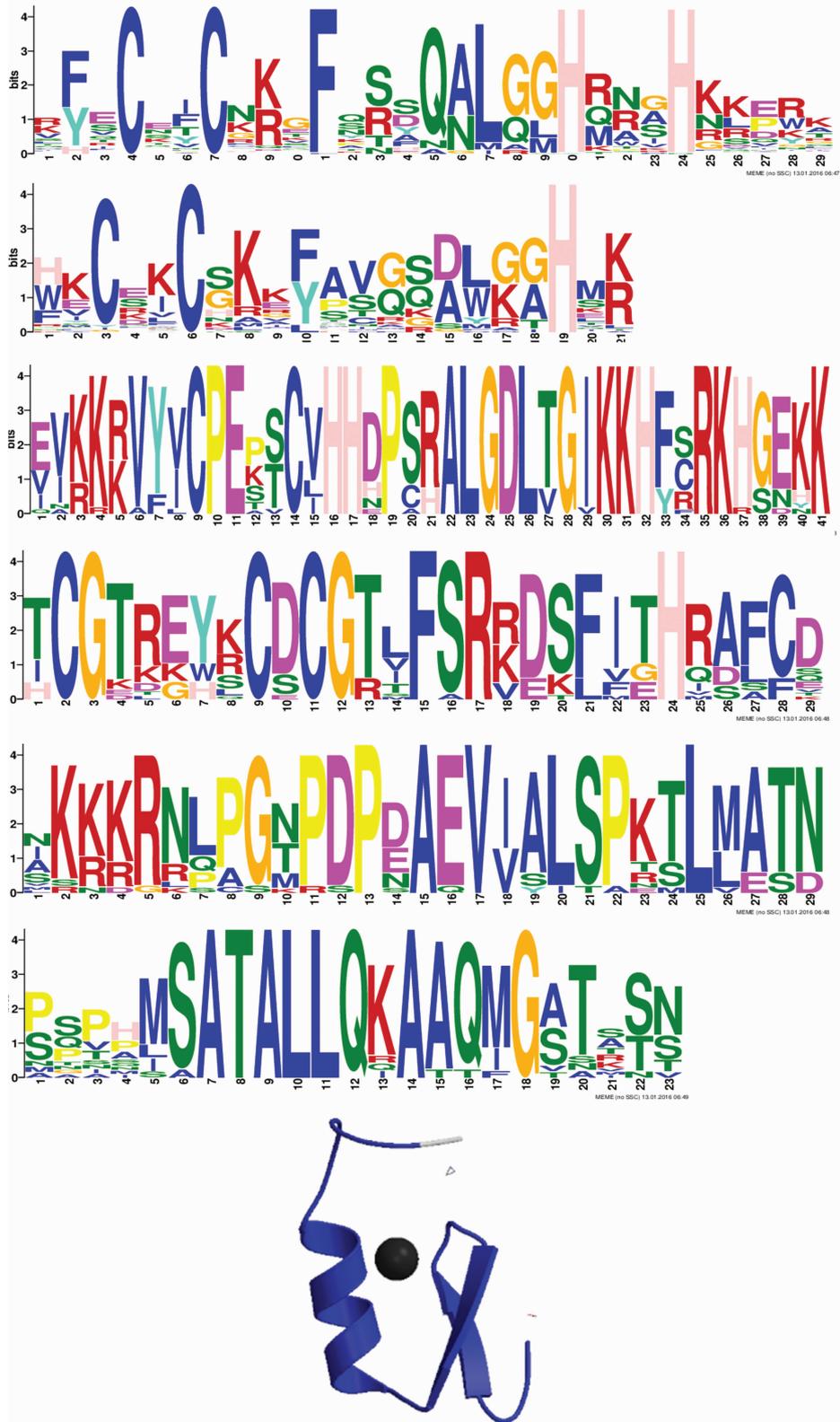


图2 番茄 C2H2 锌指蛋白保守结构域分析

DNA 序列具有重要作用,表明植物 C2H2 锌指蛋白家族可能控制植物中一些所特有的生命过程^[17]。

对 C2H2 基因家族所有成员染色体定位分析发现,C2H2 基因家族成员分布较广泛,同一亚家族成员多分散于不同的

染色体上,这也从侧面说明了 C2H2 基因家族成员作用广泛。番茄 C2H2 家族基因在番茄的根、茎、叶、花、果皮、种子等各个组织都有表达量,且表达模式多样化,同一家族的不同基因在组织特异性表达模式上存在差异。同一亚家族之间表达模

表2 番茄各个亚家族成员染色体分布情况

染色体	分布数量(个)					合计
	亚家族 I	亚家族 II	亚家族 III	亚家族 IV	亚家族 V	
1号	5	4	0	1	3	12
2号	0	0	3	3	2	8
3号	2	2	0	0	3	7
4号	1	2	0	0	4	7
5号	3	5	0	0	2	10
6号	4	4	0	0	5	13
7号	1	1	0	0	1	3
8号	1	2	0	1	1	5
9号	5	1	0	0	3	9
10号	5	0	0	0	1	6
11号	2	3	0	0	4	9
12号	0	1	0	0	0	1
0号	1	0	0	0	0	1
合计	30	25	3	5	29	92

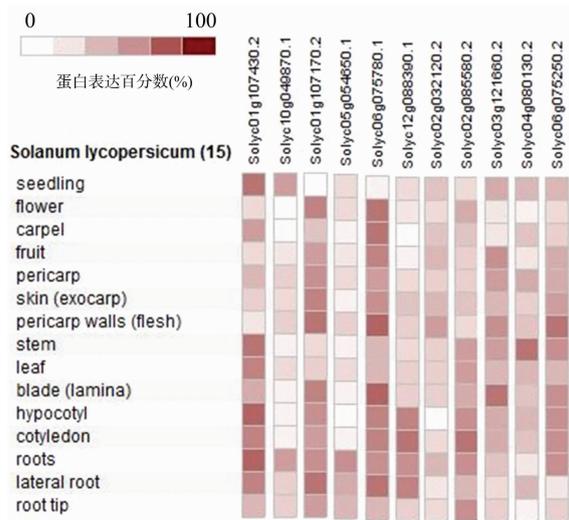


图3 番茄 C2H2 锌指蛋白组织表达模式

式也不相同,各个成员在不同组织中的表达情况也存在明显的差异,这表明各基因的表达具有一定的组织特异性,猜测这是基因的表达模式受到基因型和环境因素的共同调控的结果。该基因家族成员广泛参与了番茄的多种生物学进程,但具体生物学功能还需要分子生物学实验验证。

4 结论

从番茄全基因组中一共鉴定出了 92 条番茄 C2H2 型锌指蛋白转录因子,氨基酸序列分析结果表明,所有成员均含有 C2H2 保守结构域。进化树分析结果表明,番茄 C2H2 型锌指蛋白成员可以分为 5 个亚家族,只有 II 亚家族含有 2 个 C2H2 保守结构域。与其他植物一样,所有的 C2H2 型锌指蛋白保守结构域都具有其典型的保守序列。番茄 C2H2 型锌指蛋白

成员在染色体分布不均匀,最多的达到 13 个,最少的只有 1 个。各个成员在不同组织中的表达情况也存在明显的差异,这表明各基因的表达具有一定的组织特异性。

参考文献:

- [1] Takatsuji H. Zinc - finger proteins; the classical zinc finger emerges in contemporary plant science[J]. *Plant Molecular Biology*, 1999, 39(6): 1073 - 1078.
- [2] Pabo C O, Peisach E, Grant R A. Design and selection of novel Cys2His2 zinc finger proteins[J]. *Annual Review of Biochemistry*, 2001, 70(70): 313 - 340.
- [3] Sakamoto H, Maruyama K, Sakuma Y, et al. Arabidopsis Cys2/his2 - type zinc - finger proteins function as transcription repressors under drought, cold, and high - salinity stress conditions [J]. *Plant Physiology*, 2004, 136(1): 2734 - 2746.
- [4] Wolfe S A, Nekludova L, Pabo C O. DNA recognition by Cys2His2 zinc finger proteins [J]. *Annual Review of Biophysics and Biomolecular Structure*, 2000, 29(1): 183 - 212.
- [5] Bowman J L, Sakai H, Jack T, et al. Superman, a regulator of floral homeotic genes in *Arabidopsis* [J]. *Development*, 1992, 114(3): 599 - 615.
- [6] Sakamoto H, Araki T, Meshi T, et al. Expression of a subset of the *Arabidopsis* Cys(2)/His(2) - type zinc - finger protein gene family under water stress[J]. *Gene*, 2000, 248(1/2): 23 - 32.
- [7] 杨明磊, 晁江涛, 王大伟, 等. 烟草 C2H2 锌指蛋白转录因子家族成员的鉴定与表达分析[J]. *遗传*, 2016, 38(4): 337 - 349.
- [8] 张立慧, 王志敏, 郭航, 等. 番茄基因组学研究进展[J]. *园艺学报*, 2014, 41(9): 1802 - 1810.
- [9] 李晓旭, 刘成, 李伟, 等. 番茄 WOX 转录因子家族的鉴定及其进化, 表达分析[J]. *遗传*, 2016, 38(5): 444 - 460.
- [10] 阮美颖, 万红建, 叶青静, 等. 番茄查尔酮合成酶基因的鉴定及生物信息学分析[J]. *分子植物育种*, 2013, 11(3): 379 - 384.
- [11] Artimo P, Jonnalagedda M, Arnold K A, et al. ExPASy: SIB bioinformatics resource portal[J]. *Nucleic Acids Research*, 2012, 40(W1): W597 - W603.
- [12] Edgar R C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity[J]. *BMC Bioinformatics*, 2004, 5(1): 113.
- [13] Bailey T L, Boden M, Buske F A, et al. MEME SUITE: tools for motif discovery and searching[J]. *Nucleic Acids Research*, 2009, 37(Suppl 2): W202 - W208.
- [14] 晁江涛, 孔英珍, 王倩, 等. MapGene2Chrom 基于 Perl 和 SVG 语言绘制基因物理图谱[J]. *遗传*, 2015, 37(1): 91 - 97.
- [15] 宋冰, 洪洋, 王丕武, 等. 植物 C2H2 型锌指蛋白的研究进展[J]. *基因组学与应用生物学*, 2010, 29(6): 1133 - 1141.
- [16] 黄骥, 王建飞, 张红生. 植物 C2H2 型锌指蛋白的结构与功能[J]. *遗传*, 2004, 26(3): 414 - 418.
- [17] Takatsuji H. Zinc - finger transcription factors in plants [J]. *Cellular and Molecular Life Sciences*, 1998, 54(6): 582 - 596.