

大豆 TCP 转录因子家族结构域分析及功能预测

刘 洋^{1,2}, 张 慧^{1,2}, 辛大伟^{1,2}, 王琳琳^{1,2}, 张丽伟^{1,2}, 刘春燕², 陈庆山^{1,2}, 胡国华^{1,2}

(1. 东北农业大学 农学院, 黑龙江 哈尔滨 150030; 2. 黑龙江省农垦科研育种中心, 黑龙江 哈尔滨 150090)

摘要:利用已知的拟南芥 TCP 基因和大豆基因组数据库,通过生物信息学方法,鉴定并获得了大豆 TCP 家族基因序列,基因定位信息。共鉴定大豆 TCP 基因 54 个,分布在大豆除 14 号染色体外的其他 19 条染色体上。对 54 个基因进行进化和聚类分析及功能结构域的分析,发现全部具有高度保守的 TCP 结构域。根据结构域差异和系统发育分析的结果,将大豆 TCP 转录因子家族分成 2 个 TCP 亚家族。以生物信息学手段和已有的其他物种的 TCP 基因进行了比较分析,大豆 TCP 基因占总数的 3.5%,与拟南芥同源基因比较,序列长度存在相似性。对启动子元件分析显示,TCP 基因含有大量与在子叶、根、茎处大量表达有关的元件及对分生组织有作用的元件。

关键词:大豆;TCP 转录因子;结构域分析;功能预测

中图分类号:S565.1

文献标识码:A

文章编号:1000-9841(2012)05-0707-07

Domain Analysis and Function Prediction of TCP Transcription Factors Family in Soybean

LIU Yang^{1,2}, ZHANG Hui^{1,2}, XIN Da-wei^{1,2}, WANG Lin-lin^{1,2}, ZHANG Li-wei^{1,2}, LIU Chun-yan², CHEN Qing-shan^{1,2}, HU Guo-hua²

(1. College of Agriculture, Northeast Agricultural University, Harbin 150030, Heilongjiang; 2. The Crop Research and Breeding Center of Land-Reclamation, Harbin 150090, Heilongjiang, China)

Abstract: TCP transcription factors gene sequences were identified based on soybean genome database and *Arabidopsis* TCP genes by the bioinformatics method. Fifty-four soybean TCP genes all contained highly conservative TCP structure domain, were found and distributed in 19 chromosomes except chromosome No. 14. They could be classified into 2 sub-families according to their structure and phylogeny relationship. By compare to TCP genes of other species, the evolution structure was also be analyzed. Soybean TCP genes account for 3.5% of the total amount, and compared with *Arabidopsis* homologous genes, the sequence length exist similarities. Analysis of promoter elements showed that the TCP gene contains elements, they refer to expression in the cotyledons, roots, stems and has the function of meristem tissue. All results provide a basis for the further study of the soybean TCP transcription factor.

Key words: Soybean; TCP-domain; Structure domain analysis; Function prediction

TCP 家族在 1999 年被确认,家族成员都含有 TCP 保守结构域并编码结构相似的蛋白,其中由 59 个氨基酸组成的 bHLH 结构是结合 DNA 和蛋白互作所必需的。TCP 的名字来自最初确定的 4 个编码蛋白的基因:玉米中的 *tb1* 基因,金鱼草中的 *cyc* 基因和水稻中的 *Pcf1*, *Pcf2* 基因^[1]。玉米中的 *tb1* 基因对植物花期的顶端优势和雄花形成有重要作用^[2]。金鱼草中的 *cyc* 基因与植物两侧对称的花序结构有关^[3]。*Pcf1* 和 *Pcf2* 基因功能与水稻中 *PCAN* 基因的启动子结合有关^[4]。这 4 个基因编码的蛋白都含有一段 59 个氨基酸编码的保守结构域,即 TCP-domain。TCP 转录因子按其功能主要分为

TCP-C 和 TCP-P 2 个亚家族,在高等植物中有独特的功能,在植物发育阶段起到重要的调控作用,特别对花器官的发育和花形态建成有重要影响,2003 年首次证明了在植物中存在与花序结构调整和发育有关的 TCP-domain 家族成员。研究表明拟南芥 TCP 家族基因 (*At1g30210*) 在拟南芥花组织中有大量表达^[5-6],在棉花中 *GhTCP1* 基因在棉花的侧芽有特异表达^[7]。但对大豆 TCP 基因研究甚少,功能也尚不明确。本研究以 TCP 基因结构分析入手,分析大豆 TCP 基因家族的结构和表达特征,对大豆 TCP 转录因子在植物发育阶段的功能研究具有重要意义。

收稿日期:2012-04-19

基金项目:抗除草剂转基因大豆新品种培育项目(2011ZX08004-001-06);现代农业产业体系资助项目(CARS-04-02A);黑龙江省普通高等学校新世纪优秀人才培养计划(1252-NCET-004)。

第一作者简介:刘洋(1987-),女,在读硕士,研究方向为植物遗传育种。E-mail:yangyang44815@163.com。

通讯作者:陈庆山(1973-),男,博士,教授,主要从事大豆生物技术研究工作。E-mail:qshchen@126.com。

1 材料与方法

1.1 材料

以大豆(*Glycine max*)为研究对象,大豆全基因组数据、重复制信息和 cDNA 数据均下载于(<http://www.phytozome.net/index.php>)。拟南芥、水

稻和棉花的 TCP 基因序列和蛋白序列分别下载自 TAIR (<http://arabidopsis.org>), TIGR (<http://rice.plantbiology.msu.edu>) 和 NCBI (<http://www.ncbi.nlm.nih.gov/>) 中的 UniGene 以及 PlantTFDB (<http://plantfdb.cbi.edu.cn/>)。基因芯片数据下载于 Soybase(<http://www.soybase.org/>)(表 1)。

表 1 4 种主要作物的转录因子信息

Table 1 The transcription factor information of four main crops

数据来源 Date source	物种 Specie	TCP 转录因子数 TFs	主页 Homepage
PlnTFDB	<i>Arabidopsis thaliana</i>	28	http://plantfdb.cbi.edu.cn
PlantTFDB	<i>Gossypium hirsutum</i>	23	http://plantfdb.cbi.edu.cn
PlantTFDB	<i>Oryza sativa</i>	23	http://plantfdb.cbi.edu.cn
PlantTFDB	<i>Glycine max</i>	54	http://plantfdb.cbi.edu.cn

1.2 方法

1.2.1 TCP 基因的鉴定 利用已知的拟南芥 TCP 基因序列,通过在线工具(<http://www.phytozome.net/index.php>)获得大豆 TCP 相似基因,使用 Pfam 数据库工具对其结构域进行分析,利用 SMART(<http://smart.embl-heidelberg.de/>)在线工具比对,找出含有典型 TCP 结构域的基因,最终获得大豆 TCP 相关基因。再根据保守结构域的差异将所得 TCP 成员分成不同的亚家族。

1.2.2 多序列联配和构建系统发育树及物种分布统计 下载拟南芥、水稻和预测出的大豆 TCP 蛋白序列集,利用 Clustal X 对其进行多序列联配分析,序列联配结果使用 MEGA 4.0(<http://megasoftware.net>)程序生成 TCP 基因的系统进化树。进化树生成采用邻接法,bootstrap 值设为 500。结构域信息搜索 Pfam 25.0 数据库,统计各个物种中 TCP 基因的序列分布。

1.2.3 TCP 基因的启动子分析 通过大豆基因组浏览器(<http://www.phytozome.net/index.php>),取已鉴定的大豆 TCP 基因的开放读码框(CDSs)的 5'

端以上的 2 000 bp 序列作为启动子分析区,利用植物顺式作用元件数据库 PLACE26.0(<http://www.dna.affrc.go.jp/PLACE/signalscan.html>)分析启动子中含有的 DNA 调节元件。

1.2.4 TCP 转录因子 DNA 结合结构域及功能预测分析 对 TCP 基因进行多序列连配比较获得 TCP-domain 结构信息。利用在线工具(<http://gsds.cbi.pku.edu.cn/>)得到基因外显子-内含子结构图。NCBI 比对获得拟南芥中与大豆 TCP 基因的同源序列注释,包括片段长度,组织表达特异性。

2 结果与分析

2.1 TCP 转录因子的确定

通过已知拟南芥、棉花和水稻的 TCP 序列比对,得到 TCP-domain,应用 <http://www.phytozome.net/index.php> 获得大豆中与 TCP 相似基因 58 个,与 PlantTFDB 上下载的 TCP 转录因子比对,SMART(<http://smart.embl-heidelberg.de/>)在线工具比对,最终确定具有 TCP 保守结构域的转录因子 54 个,分布在除 14 号染色体外的 19 条染色体上(图 1)。

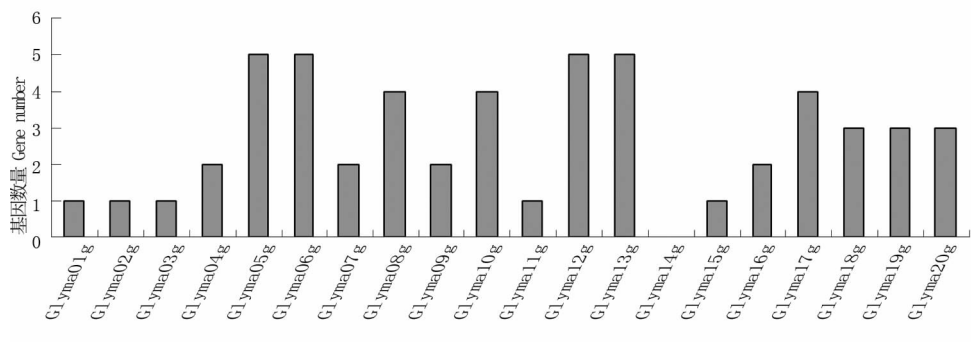


图 1 54 个 TCP 转录因子在染色体上的分布

Fig. 1 Distribution of 54 transcription factor in soybean chromosomes

同时根据所含结构域的不同可以将所得的 TCP 基因分成 TCP-C 和 TCP-P 2 个亚家族(表 2)。TCP-P 亚家族成员是类 *PCF1* 和 *PCF2* 基因,TCP-C 亚家族成员是类 *CYC* 和 *TB1* 基因。大多数的 TCP 家族成员都属于 TCP-P 亚家族,它们都含有 TCP domain,由 59 个氨基酸残基组成,可形成 bHLH 结构

域,但 TCP-C 亚家族除了 bHLH 结构域外,还有一个保守的 R 结构域。因为 *CYC* 和 *TB1* 都含有一个保守的 R 结构域,可形成一个亲水的 α 螺旋,其中极性氨基酸较多,bHLH 结构和 R 结构域会形成一种类似亮氨酸拉链形成的卷曲螺旋结构,而 *PCF1* 和 *PCF2* 中无此类结构。

表 2 转录因子定位及所属亚家族

Table 2 The positioning and subfamily of 54 transcription factors

转录因子 TF ID	TCP 结构域 TCP domain	亚家族 Subfamily	转录因子 TF ID	TCP 结构域 TCP domain	亚家族 Subfamily
Gma016094	TCP subgroup CYC/TB1 R domain	TCP-C	Gma027992	TCP subgroup	TCP-P
Gma020863	TCP subgroup CYC/TB1 R domain	TCP-C	Gma032582	TCP subgroup	TCP-P
Gma023515	TCP subgroup CYC/TB1 R domain	TCP-C	Gma033764	TCP subgroup	TCP-P
Gma025265	TCP subgroup CYC/TB1 R domain	TCP-C	Gma034498	TCP subgroup	TCP-P
Gma000351	TCP subgroup CYC/TB1 R domain	TCP-C	Gma036101	TCP subgroup	TCP-P
Gma030525	TCP subgroup CYC/TB1 R domain	TCP-C	Gma036823	TCP subgroup	TCP-P
Gma031707	TCP subgroup CYC/TB1 R domain	TCP-C	Gma038444	TCP subgroup	TCP-P
Gma038950	TCP subgroup CYC/TB1 R domain	TCP-C	Gma038799	TCP subgroup	TCP-P
Gma039271	TCP subgroup CYC/TB1 R domain	TCP-C	Gma040514	TCP subgroup	TCP-P
Gma041582	TCP subgroup CYC/TB1 R domain	TCP-C	Gma043131	TCP subgroup	TCP-P
Gma043101	TCP subgroup CYC/TB1 R domain	TCP-C	Gma043283	TCP subgroup	TCP-P
Gma044757	TCP subgroup CYC/TB1 R domain	TCP-C	Gma044797	TCP subgroup	TCP-P
Gma044882	TCP subgroup CYC/TB1 R domain	TCP-C	Gma045025	TCP subgroup	TCP-P
Gma000737	TCP subgroup	TCP-P	Gma045207	TCP subgroup	TCP-P
Gma001276	TCP subgroup	TCP-P	Gma027992	TCP subgroup	TCP-P
Gma001883	TCP subgroup	TCP-P	Gma032582	TCP subgroup	TCP-P
Gma004931	TCP subgroup	TCP-P	Gma033764	TCP subgroup	TCP-P
Gma005865	TCP subgroup	TCP-P	Gma034498	TCP subgroup	TCP-P
Gma006912	TCP subgroup	TCP-P	Gma024482	TCP subgroup	TCP-P
Gma007390	TCP subgroup	TCP-P	Gma025341	TCP subgroup	TCP-P
Gma008050	TCP subgroup	TCP-P	Gma027041	TCP subgroup	TCP-P
Gma008075	TCP subgroup	TCP-P	Gma027663	TCP subgroup	TCP-P
Gma010582	TCP subgroup	TCP-P	Gma017857	TCP subgroup	TCP-P
Gma010886	TCP subgroup	TCP-P	Gma018161	TCP subgroup	TCP-P
Gma012845	TCP subgroup	TCP-P	Gma018166	TCP subgroup	TCP-P
Gma013114	TCP subgroup	TCP-P	Gma018503	TCP subgroup	TCP-P
Gma013956	TCP subgroup	TCP-P	Gma018581	TCP subgroup	TCP-P
Gma015300	TCP subgroup	TCP-P	Gma020388	TCP subgroup	TCP-P
Gma017808	TCP subgroup	TCP-P	Gma022921	TCP subgroup	TCP-P

2.2 TCP 转录因子结构域分析

对大豆 54 个 TCP 成员进行多序列联配分析,结果表明 TCP 转录因子基因编码的蛋白序列上都有一段由 60 个左右氨基酸组成的保守结构域,称为 TCP-domain,其中碱性区域大约由 21 个氨基酸残基组成,可以形成 b-HLH 结构,含有 1 个可能的

核定位信号(NLS),提示 TCP 结构中的碱性区域可使这些蛋白定位在细胞核内,大部分 NLS 结构域处于 DNA 结合结构域的边缘或重叠部分,与其定位功能有关^[8]。而对于 NLS 结构,2 个亚家族也存在不同,TCP-C 亚家族在碱性区含有完整的 NLS 结构,而 TCP-P 亚家族只含有 NLS 的部分序列,*PCF1* 和

PCF2 氨基酸序列 N 端含有 68 个氨基酸组成的保守序列, 同源性可达 70%, 而其他区域则无同源性, 对保守区域的 5' 和 3' 端进行删除检测, 发现 bHLH 结构域对 DNA 特异结合是必需的, 缺失一小段序列就会导致功能的丧失^[1]。TCP 基因的 2 个螺旋区, 第一个是由保守的疏水残基和亲水残基组成的两亲结构, 第二个螺旋结构区含有 1 个 LXXLL 的结构。TCP domain 的 b-HLH 结构可能通过双亲螺旋的疏水表面来调节蛋白与蛋白间的相互作用, 形成同源或异源二聚体, 这与 MADS-domain 双亲螺旋 (K 区) 调节机理很相似^[1]。例如在 PCF 中, TCP 结构域对于同源或异源二聚体的形成也是必需的。经在线工具 SMART (<http://smart.embl-heidelberg.de/>)^[9] 比对可知保守结构域可位于 ATG 起始密码子较近的地方或是在序列中部位置, 还可能在离 TAG 终止密码子较近的区域, 说明 TCP 保守结构域编码序列的位置并不唯一。内含子外显子结构分析^[10] 表明, 54 个基因的内含子和外显子特征也不尽相同, 其中 30 个基因不含有内含子, 19 个基因有 1~3 个内含子, 5 个基因有 4~5 个内含子, 而同一染色体的基因所含内含子数相似。

2.3 TCP 基因物种分布和系统发育分析

TCP 基因在各物种中的分布由在线软件 (<http://pfam.sanger.ac.uk>) 获得 (表 3), TCP 基因分布在 24 个目中, 共 1 532 个基因, 多数存在于高等植物中, 基本覆盖双子叶植物, 少数存在未知生物中。TCP 基因在物种中分布不均, 在 6 个主要目中存在较多, 占总数的 86.88%, 其余分布在其他 18 个目中, 含量 1~40 个不等, 其中有 10 个基因存在于还不确定的物种中。TCP 基因在唇形目和豆目中含量最高。其中豆目含有 220 个基因, 大豆 TCP 基因 54 个, 占豆目的 24.5%, 占 TCP 总数的 3.5%。

通过 ClustalX 对其进行多序列联配分析, 序列联配结果使用 MEGA 4.0 (<http://megasoftware.net>) 程序生成 TCP 基因的系统进化树, 将 54 个 TCP 基因分成 2 个大组, 7 个小组, 每个小组有 5~11 个基因不等, 亲缘关系紧密但又有明显的差异, 2 个亚家族中, TCP-P 家族又可分成 6 个小组, 亚家族分类和系统发生树进行比较可发现, 并非所有的 TCP-C 成员都被分成了一类, TCP-P 和 TCP-C 成员相互穿插, 序列比较发现, 系统发生树结果与序列长度有一定关系, 但并非一一对应, 说明 2 个亚家族在进化过程初期有相同的保守序列和功能, 在进化过程中由于选择压力的不同, 形成或退化了完整或部分的其他结构域, 行使不同或相似的功能, 在序列上

表 3 TCP 转录因子在物种中的分布
Table 3 TCP transcription factors in the distribution of species

所属目 Subordinate orders	含 TCP 基因物种数 Including TCP gene species	所含 TCP 基因序列数 Contains TCP number
葡萄目 <i>Vitis vinifera</i>	1	22
蔷薇目 <i>Rosa multiflora</i> Thunb	1	11
无患子 <i>Sapindus mulorossi</i> Gaertn	1	1
茄目 <i>Fructus Lycopersici</i> Esculenti	8	25
虎耳草 <i>Saxifraga stolonifera</i> Curt	2	3
天门冬目 <i>Asparagales</i>	2	2
酢浆草目 <i>O. corniculata</i> L	1	2
锦葵目 <i>Malva sinensis</i> Cavan	2	3
百合目 <i>Lilium</i>	1	1
菊目 <i>Matricaria</i>	8	34
卷柏目 <i>Selaginellales</i>	1	9
松柏目 <i>Poria cocos</i>	12	29
毛茛目 <i>Ranunculales</i>	1	1
川绿断目 <i>Rosaceae</i>	29	91
姜目 <i>Zingiber officinale</i> Rosc	1	1
无分类真核目	40	176
龙胆目 <i>Gentiana sino-ornata</i>	1	4
金虎尾目 <i>Malpighia</i>	42	143
白花菜目 <i>Orychophragmus violaceus</i>	4	35
禾本科 <i>Echinochloa crusgalli</i>	43	207
石竹目 <i>Caryophyllaceae</i>	2	8
豆目 <i>Arachis hypogaea</i>	119	220
唇形目 <i>Lamiales</i>	222	494
未确定 Undetermined	2	10
总计 Total	546	1532

有保守性, 在功能上也具有重复性。对大豆和拟南芥的序列结构比较分析, 获得与大豆 TCP 基因同源的拟南芥基因信息及 TCP 基因大量表达的组织 (表 4)。同源比对结果显示, 同源性与序列长度存在一定的关系, 同源基因的序列长度相似, 但也有个例差异较大。同源 TCP 家族基因在大豆和拟南芥中序列长度相似, 还有共同的 TCP 保守结构域, 转录后蛋白结构和序列相似, 说明该类蛋白在进化过程具有保守性。同源序列分布在不同的染色体上, 根据已知的拟南芥同源序列的组织表达及功能, 拟南芥 TCP 基因主要在花、芽、叶、种子、植物根系、营养和分生组织中表达, 可预测相应大豆 TCP 基因在植物中的表达特异性及对顶端花序形态的作用。

表 4 大豆 TCP 同源基因与拟南芥同源基因的的序列比较

Table 4 *Glycine max* TCP homologous genes and *Arabidopsis* gene sequence of the same

同源基因 Homologous gene	所在染色体 Chromosome	序列长度 Sequence length/bp	拟南芥同源基因 <i>Arabidopsis</i> homologous genes	序列长度 Sequence length/bp	拟南芥中表达组织 Express organization
Gma001883	04	500			长角果
Gma010886	06	1057	Ath004585	567	
Gma006912	13	447	Ath030174	1083	
Gma018166	10	399			
Gma000351	13	566			
Gma016094	18	849	Ath016115	1080	花
Gma043101	08	540	Ath018589	1071	
Gma044757	19	595			
Gma038444	07	705			
Gma005865	18	744			
Gma000737	19	800	Ath006895	1068	芽,叶,种子
Gma017857	08	891			
Gma007390	16	897			花分生组织,根叶
Gma036823	12	1143	Ath005596	1026	营养组织
Gma008075	07	851			花分生组织,根叶
Gma013956	03	822	Ath008676	1071	营养组织
Gma020863	06	837			
Gma023515	04	834	Ath026854	1302	子叶,茎
Gma039271	17	726			
Gma041582	05	741			
Gma025265	15	1221			花分生组织,根叶
Gma030525	13	1599	Ath021099	1098	营养组织
Gma031707	08	1395			
Gma036101	02	1090			芽,花分生组织
Gma012845	17	1119	Ath026943	1470	根,叶
Gma010582	01	1531			营养组织

2.4 大豆 TCP 基因家族启动子分析

对上游 2 000 bp 启动子区域进行分析,得到的 54 TCP 基因的启动子主要含有 ABRE, ACGT, ARR1, DRE, LTRF, MYB, MYC, NOD, OSE, SOR, WRKY 元件(表 5)。其中,所有的 TCP 基因都含有多个 MYB 和 MYC 元件,MYB 和 MYC 元件都属于典型的转录因子调节元件。与许多转录因子不同的是 TCP 家族成员只有一部分含有 ABRE(ABA-responsive element,脱落酸响应元件),DRE(dehydra-

tion-responsive element,脱水反应元件),LTRE(low-temperature responsive element,低温应答元件)元件,分别占总数的 72%,35% 和 53%,而且含量很少。此外,90% 以上的基因都含有 NOD,OSE,SOR 元件,NOD 元件是典型的大豆基因的启动子元件。95% 以上的基因都含有 ACGT-box 和 ARR1 元件,而 53 个基因都含有 WRKY 元件。只有 15% 的成员含有全部 11 类元件,这 15% 的基因会在植物发育阶段起到关键作用。

表 5 大豆 TCP 基因启动子中 NOD OSE SOR MYB 和 MYC 等元件的数量

Table 5 The number of NOD SOR OSE MYB and MYC promoter elements in soybean genes

基因 Gene	元件 Promoter elements										
	ABRE	ACGT	ARR1	DRE	LTRE	MYB	MYC	NOD	OSE	SOR	ERKY
Glyma18g58	0	2	3	0	0	10	15	2	2	2	8
Glyma06g21	3	1	8	1	1	8	7	2	3	11	9
Glyma04g26	0	2	4	1	0	4	14	4	3	0	4
Glyma15g01	3	12	12	1	2	14	8	4	4	3	4
Glyma13g08	0	3	5	0	0	4	15	4	3	1	3

续表 5

基因 Gene	元件 Promoter elements										
	ABRE	ACGT	ARR1	DRE	LTRE	MYB	MYC	NOD	OSE	SOR	ERKY
Glyma13g26	0	1	10	0	0	18	25	9	10	3	27
Glyma08g15	0	2	2	0	0	13	4	4	4	1	11
Glyma10g34	1	2	4	0	0	16	15	4	4	3	5
Glyma17g16	1	3	2	0	0	14	15	5	7	2	5
Glyma05g03	0	1	3	0	1	11	7	6	5	2	11
Glyma08g29	0	1	3	0	0	14	12	3	3	3	6
Glyma19g01	0	2	5	0	0	14	8	4	4	1	7
Glyma05g27	0	1	4	0	1	14	6	5	5	3	15
Glyma19g01	1	2	7	0	0	18	11	3	4	3	10
Glyma05g00	4	0	8	1	3	13	7	8	5	3	3
Glyma04g34	0	0	8	0	0	11	8	2	5	3	4
Glyma20g23	1	2	4	0	3	15	14	5	5	4	12
Glyma18g57	3	3	4	0	0	12	7	4	4	1	5
Glyma13g21	1	1	0	1	1	12	12	0	0	3	0
Glyma16g08	3	2	3	1	2	17	4	3	3	4	5
Glyma20g20	3	2	1	1	3	14	4	10	0	5	2
Glyma07g01	3	3	4	0	3	19	13	2	2	5	6
Glyma01g01	5	4	6	0	1	13	13	8	2	4	9
Glyma06g24	3	4	5	2	2	15	6	1	2	4	5
Glyma17g16	2	2	2	0	1	13	8	2	0	4	4
Glyma10g49	3	2	13	2	4	17	8	6	5	5	12
Glyma03g09	5	8	7	0	2	18	15	2	2	6	9
Glyma06g26	3	5	4	2	2	15	6	1	2	4	5
Glyma10g31	2	3	3	1	3	15	6	2	2	3	5
Glyma08g21	1	3	7	0	0	12	12	3	4	1	7
Glyma12g34	0	1	4	0	0	13	14	2	2	2	6
Glyma10g01	1	1	2	0	0	16	5	4	4	3	5
Glyma06g33	4	1	7	0	1	14	16	2	3	7	7
Glyma06g42	6	2	8	0	0	13	12	3	1	2	15
Glyma17g06	1	3	2	0	0	14	15	5	7	2	4
Glyma19g26	0	2	7	0	0	16	20	5	4	9	8
Glyma08g47	6	3	3	0	0	10	15	6	5	5	12
Glyma12g32	2	4	8	2	1	11	16	0	0	1	7
Glyma11g20	0	2	0	0	0	12	10	1	1	3	0
Glyma12g10	1	2	10	0	0	10	7	4	3	5	12
Glyma18g11	5	3	3	0	2	13	14	1	1	3	4
Glyma17g16	2	3	2	0	1	13	5	4	7	2	5
Glyma13g34	3	2	11	0	0	17	8	4	4	2	10
Glyma13g39	2	6	4	2	2	10	12	0	0	1	3
Glyma02g18	5	3	8	0	1	10	16	0		2	4
Glyma12g27	3	2	6	0	2	13	8	4	7	6	6
Glyma07g23	2	2	1	3	3	12	12	0	0	3	1
Glyma12g26	1	4	16	0	0	10	7	4	3	5	12
Glyma16g04	5	1	14	0	0	13	15	2	2	8	3
Glyma05g03	3	2	6	2	2	8	8	2	2	6	2
Glyma20g05	0	0	6	1	4	6	14	0	0	4	6
Glyma09g42	1	3	8	1	3	15	7	0	0	5	4
Glyma09g44	1	0	9	3	0	7	14	10	0	6	4
Glyma05g01	0	3	4	4	1	13	11	6	4	4	8

3 讨 论

TCP 转录因子家族在植物发育阶段对器官的背腹极性建立的关键作用已被广泛认可。金鱼草中的 TCP 家族基因 *cyc* 与 *dich* 对花的发育影响很大^[11-12]。玉米中 *tb1* 基因有抑制侧枝生长和雄花形成的作用。水稻中 PCF1 和 PCF2 转录因子是促使分生组织中 PCNA 启动子的组织特异表达所必需的^[13]。本研究获得的 54 个大豆 TCP 家族成员,大量多于已有研究的拟南芥、棉花、水稻等作物,因此 TCP 基因在大豆发育过程中有更广泛的作用。同源和进化分析表明,TCP 基因可分为 7 个小类,说明在进化过程中,基因在结构和表达上都有不同程度的变化,使其功能在不同的植物上既有共性又有独特性。大豆 TCP-C 亚家族成员较少,但都属于类 *CYC* 和 *Tb1* 基因,因此可预测 TCP-C 亚家族基因可能对大豆花序形态建成有作用。系统进化树与拟南芥 TCP 基因同源序列的获得,得到 TCP 基因的进化和亲缘关系,有助于推测 TCP 基因在植物组织器官的表达性和基因功能。TCP-C 亚家族基因主要存在于植物的分生组织中,表达特异性较强,而 TCP-P 亚家族基因表达组织较广泛,除在含有分裂细胞多的组织中有大量表达外,在花、子叶、根、种子和一些营养器官也有表达。推测 TCP 家族基因可能作用于整个植物发育阶段。

TCP 基因物种分布数据显示,TCP 基因主要分布在唇形目,禾本目和豆目中,豆目有 220 个 TCP 基因,占总数的 14.4%,其中大豆有 54 个 TCP 基因,占 TCP 基因总数的 3.5%,大豆 TCP 基因明显高于其他双子叶植物,可能与进化过程中发生的 2 次基因复制有关,大豆 TCP 基因进化既有保守性又有特异性,其功能不仅对花序结构和营养器官发育有影响,也可能广泛作用在植物发育的各个阶段,应该还有其他的调控机制和功能。启动子分析结果表明,54 个基因都含有 MYB、MYC 转录调控元件,但 ABRE、DRE 和 LTRE 含量较少,ABRE、DRE 和 LTRE 是植物逆境胁迫中发挥重要作用的 3 类元件^[14],所以 TCP 转录因子对植物抗逆境胁迫的影响不会太大。TCP 家族基因普遍含有 NOD、OSE、SOR、ACGT-box 和 ARR1 元件。NOD 元件是大豆启动子元件,OSE 元件对调控启动子活性及豆科根瘤的形成有重要关系,SOR 元件对子叶、根和茎中基因的大量表达起关键作用。ARR1 元件对调节细胞分裂有重要作用,ACGT-box 在种子发育阶段起到关键作用。这些都进一步说明 TCP 家族基因在植物组织中的表达特异性,在种子发育和花期花序的

形态方面有重要影响,而且因为 NOD、OSE 调控元件的存在,其功能可能与大豆特异性表达有关,对顶端优势和雄花的形成有关键作用。但 *Glyma11g20* 基因的几个主要元件含量都为 0,说明它很可能有其他的功能或调控方式。对于序列结构的分析结果,54 个基因都含有 TCP 结构域,但结构域所在位置不唯一,可能在形成三维结构时通过扭曲盘绕形成保守的 TCP 结构域。TCP domain 的 b-HLH 结构可能通过双亲螺旋的疏水表面来调节蛋白与蛋白间的相互作用,这一特点与 MADS-domain 双亲螺旋的 K 区调节机理很相似^[1]。而大量实验证明 MADS-box 对拟南芥和酸浆等一些浆果类植物花序结构的形成起到关键作用^[15]。因此可推测大豆 TCP 转录因子可能与大豆花序形态的建成有关,影响大豆的侧枝数及顶端优势,进而影响大豆产量。两个亚家族相差的 R 区结构很可能与 DNA 专一性结合或蛋白与蛋白间的互作有关。大豆 TCP-C 亚家族基因的拟南芥 TCP 同源基因多在花中大量表达,特异性较强,可能与 R 域结构有关。

参考文献

- [1] Cubas P, Lauter N, Doebley J, et al. The TCP domain: a motif found in proteins regulating plant growth and development [J]. *Plant Journal*, 1999, 18(2): 215-222.
- [2] Doebley J, Stec A, Gustus C. Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance [J]. *Genetics*, 1995, 141(1): 333-346.
- [3] Luo D, Carpenter R, Vincent C, et al. Origin of floral asymmetry in *Antirrhinum* [J]. *Nature*, 1996, 383: 794-799.
- [4] Kosugi S, Ohashi Y. PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene [J]. *Plant Cell*, 1997, 9: 1607-1619.
- [5] 雷娟, 魏刚, 朱玉贤. 拟南芥 TCP 家族转录因子 At2g31070 的转录激活活性鉴定与表达谱分析 [J]. *分子植物育种*, 2005(1): 31-35. (Lei J, Wei G, Zhu Y X. *Arabidopsis* TCP family At2g31070 transcription factors of transcription activate active appraisal and express spectrum analysis [J]. *Molecular Plant Breeding*, 2005(1): 31-35.)
- [6] Ababakeli S, Wei G, Lei J, et al. Cloning and trans-activation identification of At1g30210, a new member of TCP family highly expressed in flower in *Arabidopsis* [J]. *Molecular Plant Breeding*, 2005, 1(6): 26-30.
- [7] 刘伟娜, 韩建民, 董金皋, 等. 棉花 TCP 家族转录因子基因 *GhTCP1* 的克隆与表达分析 [J]. *棉花学报*, 2010, 22(5): 409-414. (Liu W N, Han J M, Dong J G, et al. Cloning and expression analysis cotton TCP family transcription factors *GhTCP1* gene [J]. *Cotton Science*, 2010, 22(5): 409-414.)
- [8] Jiang C J, Imamoto N, Matsuki R, et al. Functional characterization of a plant importin alpha homologue. Nuclear localization signal (NLS)-selective binding and mediation of nuclear import of nls proteins in vitro [J]. *Journal of Biological Chemistry*, 1998, 273: 24083-24087.

(下转第 717 页)

- [J]. 大豆科学, 2010, 29(1): 151-156. (Zhao Y, Liu X X, Zhang Q L, et al. Advances of studies on seed-specific promoters of soybean[J]. Soybean Science, 2010, 29(1): 151-156.)
- [2] Chiera J M, Finer J J, Grabau E A. Ectopic expression of a soybean phytase in developing seeds of *Glycine max* to improve phosphorus availability [J]. Plant Molecular Biology, 2004, 56(6): 895-904.
- [3] Ding S H, Huang L Y, Wang Y D, et al. High-level expression of basic fibroblast growth factor in transgenic soybean seeds and characterization of its biological activity [J]. Biotechnology Letters, 2006, 28(12): 869-875.
- [4] Sarmiento C, Ross J H, Herman E, et al. Expression and subcellular targeting of a soybean oleosin in transgenic rapeseed. Implications for the mechanism of oil-body formation in seeds [J]. Plant Journal, 1997, 11(4): 783-796.
- [5] Philip R, Darnowski D W, Maughan P J, et al. Processing and localization of bovine β -casein expressed in transgenic soybean seeds under control of a soybean *lectin* expression cassette [J]. Plant Science, 2001, 161(2): 323-355.
- [6] Gaur V S, Singh U S, Kumar A. Transcriptional profiling and in silico analysis of Dof transcription factor gene family for understanding their regulation during seed development of rice (*Oryza sativa* L.) [J]. Molecular Biology Reports, 2011, 38(4): 2827-2848.
- [7] Jeong S C, Yang K, Park J Y, et al. Structure, expression, and mapping of two nodule-specific genes identified by mining public soybean EST databases [J]. Gene, 2006, 15(383): 71-80.
- [8] Meng Q C, Zhang C H, Gai J Y, et al. Molecular cloning, sequence characterization and tissue-specific expression of six NAC-like genes in soybean [*Glycine max* (L.) Merr.] [J]. Journal of Plant Physiology, 2007, 164(8): 1002-1012.
- [9] Chauhan H, Khurana N, Agarwal P, et al. Heat shock factors in rice (*Oryza sativa* L.): genome-wide expression analysis during reproductive development and abiotic stress [J]. Molecular Genetics Genomics, 2011, 286(2): 171-187.
- [10] Terauchi K, Asakura T, Ueda H. Plant-specific insertions in the soybean aspartic proteinases, soyAPI and soyAP2, perform different functions of vacuolar targeting [J]. Plant Physiology, 2006, 163(8): 856-862.
- [11] Yoder D W, Nampalsansuk M, Pirtle I L, et al. Molecular cloning and nucleotide sequence of a gene encoding a cotton palmitoyl-acyl carrier protein thioesterase [J]. Biochimica et Biophysica Acta-Gene Structure and Expression, 1999, 1446(3): 403-413.
- [12] Holmberg N, Harker M, Gibbard C L, et al. Sterol C-24 methyltransferase type 1 controls the flux of carbon into sterol biosynthesis in tobacco seed [J]. Plant Physiology, 2002, 130(3): 303-311.
- [13] Chatthai M, Forward B S, Yevtushenko D, et al. 2S storage protein gene of Douglas-fir: characterization and activity of promoter in transgenic tobacco seeds [J]. Plant Physiology and Biochemistry, 2004, 42(5): 417-423.
- [14] Moreno-Risueno MA, González N, Díaz I, et al. FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and *Arabidopsis* [J]. Plant Journal, 2008, 53(6): 882-894.
- [15] Yoshino M, Nagamatdu A, Tsutsumi K, et al. The regulatory function of the upstream sequence of the β -conglycinin α subunit gene in seed-specific transcription is associated with the presence of the RY sequence [J]. Genes and Genetic Systems, 2006, 81(2): 135-141.
- [16] Chung K J, Hwang S K, Hahn B S, et al. Authentic seed-specific activity of the Perilla oleosin 19 gene promoter in transgenic *Arabidopsis* [J]. Plant Cell Reports, 2008, 27(1): 29-37.
- [17] Chandrasekharan M B, Bishop K J, Hall T C. Module specific regulation of the beta-phaseolin promoter during embryogenesis [J]. Plant Journal, 2003, 33(5): 853-866.
- [18] Kim M J, Kim J K, Shin J S, et al. The SeBHLH transcription factor mediates trans-activation of the *SeFAD2* gene promoter through binding to E- and G-box elements [J]. Plant Molecular Biology, 2007, 64(4): 453-466.
- [19] Washida H, Wu C Y, Suzuki A, et al. Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene *GluB-1* [J]. Plant Molecular Biology, 1999, 40(1): 1-12.
- [20] Niu X, Adams C C, Workman J L, et al. Binding of the wheat basic leucine zipper protein EmBP-1 to nucleosomal binding sites is modulated by nucleosome positioning [J]. Plant Cell, 1996, 8: 1569-1587.
- [21] Chamberland S, Daigle N, Bernier F. The legumin boxes and the 3' part of a soybean beta-conglycinin promoter are involved in seed gene expression in transgenic tobacco plants [J]. Plant Molecular Biology, 1992, 19(6): 937-949.

(上接第 713 页)

- [9] 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(Database issue): D257-260.
- [10] 郭安源, 朱其惠, 陈新. 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. GSDS: Genetic structure the display system [J]. Genetic, 2007, 29(8): 1023-1026.)
- [11] Coen E S. Floral symmetry [J]. EMBO Journal, 1996, 15: 6777-6788.
- [12] Luo D, Carpenter R, Vincen C, et al. Origin of floral asymmetry in *Antirrhinum* [J]. Nature, 1996, 383: 794-799.
- [13] Kosugi S, Ohashi Y. PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene [J]. Plant Cell, 1997, 9: 1607-1619.
- [14] 聂丽娜, 夏兰琴, 徐兆师, 等. 植物基因启动子的克隆及其功能研究进展 [J]. 植物遗传资源学报, 2008, 9(3): 385-391. (Nie L N, Xia L Q, Xu Z S, et al. Plant gene promoter of the cloning research progress and its functions [J]. Journal of Plant Genetic Resources, 2008, 9(3): 385-391.)
- [15] He C Y, Tian Y, Saedler R, et al. The MADS-domain protein MPF1 of *Physalis xoridana* controls plant architecture, seed development and Flowering time [J]. Planta, 2010, 231: 767-777.