

4 个物种 CCT 结构域基因家族的序列进化分析*

陈华夏 申国境 王 磊 邢永忠**

华中农业大学生命科学技术学院/作物遗传改良国家重点实验室, 武汉 430070

摘要 通过分析 4 个完成测序和注释的植物基因组, 系统地分离鉴定了 97 个水稻、玉米、高粱和拟南芥的 CCT 结构域基因, 并对相应蛋白质的结构和基因之间的系统演化关系进行了分析。结果表明: CCT 结构域基因的蛋白质结构和特性在不同物种之间具有广泛的变异; 不同基因组中 CCT 结构域基因通过染色体复制扩展了基因家族成员。根据其 CCT 结构域分为 4 组, 其中 1 个亚组集中了绝大部分的禾本科 CCT 结构域基因, 该亚组基因可能特异参与了禾本科作物开花时间的调控。基因家族成员的扩展, 蛋白质结构的改变以及基因表达模式的变异共同导致了 CCT 结构域基因家族成员在物种间和物种内的功能分化。

关键词 CCT 结构域基因; 染色体复制; 蛋白质结构; 功能分化

中图分类号 S 188; Q 811.4 **文献标识码** A **文章编号** 1000-2421(2010)06-0669-08

开花是被子植物完成生命周期的重要发育事件, 正确的开花时间保证了某个物种能成功地适应特殊的地理分布和完成生命周期并繁衍后代。在过去的 20 a 中, 得益于基因组的研究和大量突变体材料的积累, 开花的分子机制在双子叶模式植物拟南芥和单子叶模式植物水稻中得到了较为深入的剖析^[1-4]。

研究表明, 在拟南芥中存在 4 条调控开花时间的途径即光调控途径、自主开花途径、春化途径以及赤霉素信号途径^[5]。对于单子叶模式植物水稻, 主要存在光周期调控的开花途径^[6]。CO (CONSTANS) 是第 1 个被克隆的参与拟南芥光调控开花的基因, 在长日照条件下促进拟南芥开花^[7]。CO 编码的蛋白质的 C 端有一段编码 43~45 个氨基酸的核定位序列, 随后该段氨基酸在 CONSTANS 类基因和 TOC1 编码氨基酸的 C 端中也被发现保守的存在, 因此被命名为 CCT 结构域^[8]。近年来的研究表明, 除了核定位功能外^[9], CCT 结构域基因与 HAP2 类基因的 NF-YA1/2 亚基具有相同的功能位点而能与 NF-YB, NF-YC 结合, 可能介导了蛋白质之间的互作^[10]。

水稻作为单子叶模式植物, 也发现了一系列含 CCT 结构域的基因参与了水稻开花(抽穗)的调控。

其中 2 个主效 QTL *Hd1* 和 *Ghd7* 均含有 CCT 结构域。与 CO 相反, *Hd1* 在长日照条件下抑制水稻开花, 而在短日照条件下则起促进作用^[11]。*Ghd7* 虽在长日照条件下抑制水稻开花, 但是在短日照条件下则不起作用^[12]。而水稻另 1 个 CCT 结构域基因 *OsCO3* 则在短日照条件下抑制水稻开花^[13]。最近, 发现 1 个 CCT 结构域基因 *OsCOL4* 在长日和短日照条件下均抑制水稻开花^[14]。上述结果表明, CCT 结构域基因虽然在不同物种中都保守地参与对开花时间的控制, 但是不同成员的具体功能在物种间和物种内都发生了很大的分化。

2003 年, Simon 等^[15]对拟南芥、水稻和大麦中分离出的 CO 类基因进行了系统的进化树分析, 而多个植物物种的基因组序列的释放则为系统分析 CO 类基因的进化提供了可能, 阐述这些基因之间的进化关系将为理解 CO 类基因功能的保守和分化提供重要的信息。本研究利用 CCT 结构域的序列检索已完成测序及注释的 4 种植物(拟南芥、水稻、玉米、高粱)的基因组获得含有 CCT 结构域基因, 利用生物信息学的方法, 对这 4 个物种中的 CCT 结构域基因构建系统发生树, 分析基因和蛋白质结构, 对 CCT 结构域基因家族的进化关系及功能异化的可能原因进行了探讨。

收稿日期: 2010-04-28; 修回日期: 2010-08-30

* 国家自然科学基金项目(30830064)资助

** 通讯作者. E-mail: yzxing@mail.hzau.edu.cn

陈华夏, 女, 1985 年生, 硕士. 研究方向: 水稻分子生物学. E-mail: chenhuaxia@webmail.hzau.edu.cn

1 材料与方法

1.1 CCT 结构域基因的分离及相应蛋白质的结构域分析

利用水稻中已发表的多效基因 *Ghd7* 编码的蛋白质中的保守 CCT 结构域氨基酸序列 EREAK-LMRYKEK RKKRCYEQIRYASRKAYAE MR-PRVRGRFAKE 作为起始序列,通过 tBLASTn 和 BLASTp 检索数据库 NCBI (<http://www.ncbi.nlm.nih.gov/>)、TIGR (<http://rice.plantbiology.msu.edu/index.shtml>)、KOME (<http://cdna01.dna.affrc.go.jp/cDNA/>)、TAIR (<http://www.arabidopsis.org/>)、MaizeGDB (<http://www.maizegdb.org/>) 和 PHYTOZOME (<http://www.phytozome.net/index.php>),以 $E < 0.1$ 为阈值,得到水稻和拟南芥中编码 CCT 结构域的基因所对应的氨基酸序列,同时获得玉米及高粱基因组中所对应的核苷酸序列。

按照植物转录因子网站 (<http://planttfdb.cbi.pku.edu.cn/>) 的注释,下载拟南芥中 *CONSTANS* 类基因的氨基酸序列,并利用 tBLASTn、BLASTp 搜索 TAIR 数据库 (<http://www.arabidopsis.org/>) 完善拟南芥中含 CCT 结构域的基因信息。玉米和高粱搜索得到的核苷酸序列通过 GENSCAN (<http://genes.mit.edu/GENSCAN.html>) 预测基因的编码方式得到相应的氨基酸序列。对初步得到的含 CCT 结构域基因,利用 SMART 网站 (<http://smart.embl-heidelberg.de/>) 和 PFAM 网站 (<http://pfam.sanger.ac.uk/>) 分析其蛋白质序列,确认含有 CCT 结构域。在水稻的 41 个基因编码的氨基酸中有 2 个 (LOC_Os11g011100、LOC_Os12g011100) 检测不到 CCT 保守结构域的存在,而植物转录因子网站 (<http://planttfdb.cbi.pku.edu.cn/>) 预测该 2 个基因为包含 CCT 结构域的基因,这 2 个基因仍作为 CCT 结构域基因纳入分析。

1.2 CCT 结构域基因的结构分析与染色体定位及复制事件的确定

根据 TIGR6.0、PLANTTFDB 和 TAIR 的注释获得水稻和拟南芥的 CCT 结构域基因的编码模式,玉米和高粱的 CCT 结构域基因则通过在线软件 GENSCAN (<http://genes.mit.edu/GENSCAN.html>) 预测相应的编码模式,获得对应的内含子和

外显子信息。

利用 TIGR 数据库 (<http://rice.plantbiology.msu.edu/index.shtml>) 中搜索水稻 CCT 结构域基因在染色体上的位置信息,从 TIGR 中下载水稻 12 条染色体的数据信息,利用 MapInspect 软件将这些基因定位于染色体上。利用 TAIR 网站在线工具 (<http://www.arabidopsis.org/jsp/Chromosome-Map/tool.jsp>) 将拟南芥 CCT 结构域基因定位于拟南芥 5 条染色体上,根据 TIGR 中提供的 2 种不同大小片段复制信息,将包含在大片段复制事件中的 CCT 结构域基因在染色体上标示出来。

1.3 CCT 结构域基因家族系统进化分析

利用软件 MEGA 4.0^[16] 中 Neighbor-Joining (N-J) 方法对 4 个物种的 CCT 结构域基因按照 CCT 保守结构域的氨基酸序列进行系统进化树的构建,同时利用 Bootstrap 的重抽样方法 (1 000 次重复) 检测各分支的稳定性及可信度。

2 结果与分析

2.1 CCT 结构域基因的分离

为了分离水稻、拟南芥、玉米和高粱的 CCT 结构域基因,利用水稻中一个控制抽穗期、株高和穗粒数的多效基因 *Ghd7* 编码的包含有 44 个氨基酸的 CCT 结构域氨基酸的序列 (EREAKLMRYKEK RKKRCYEQIRYASRKAYAE MRPRVRGRFAKE) 比对相应的基因组,在水稻和拟南芥中分别分离到 41 和 40 个含有 CCT 结构域的基因,而在玉米和高粱中,分别分离到 7 和 9 个含有 CCT 结构域的基因 (表 1)。CCT 结构域基因和编码的蛋白质之间不是一一对应的,其中水稻和拟南芥分别有 15 和 14 个 CCT 结构域基因存在选择性剪接模式,分别编码多种蛋白质,在 CCT 结构域基因系统进化分析时选择翻译后最长的一条氨基酸序列做相关分析。

2.2 CCT 结构域基因在染色体上的分布及复制事件

水稻的 CCT 结构域基因分布于全部的 12 条染色体上,其中在第 2、3、6、11 染色体上分布较为广泛,而第 1、4 染色体则只有 1 个成员分布。而拟南芥中 40 个 CCT 结构域基因也分布于全部的 5 条染色体,其中第 1 和第 5 染色体分布最为丰富。根据基因组注释信息,仅在玉米的第 2、5、9、10 染色体上分离到 7 个 CCT 结构域基因,而高粱的 9 个 CCT 结构域基因则分布于 5 条染色体上 (表 1)。

表 1 CCT 结构域基因信息¹⁾

Table 1 List of 97 CCT domain genes identified in rice, *Arabidopsis*, sorghum and maize

基因编号 Gene number	染色体 Chromosome	基因长度 Gene length/bp	蛋白质 Protein			基因编号 Gene number	染色体 Chromosome	基因长度 Gene length/bp	蛋白质 Protein		
			氨基酸长度 Amino acid length	分子质量 Molecular mass/ku	等电点 Isoelectric point				氨基酸长度 Amino acid length	分子质量 Molecular mass/ku	等电点 Isoelectric point
Os01g61900	1	1 639	336	35.93	7.28	At1g04500	1	2 934	386	42.77	4.28
Os02g01990	2	1 558	324	35.83	4.21	At1g05290	1	1 395	351	39.58	4.60
Os02g05510*	2	3 801	329	34.86	4.73	At1g07050	1	1 048	195	23.32	4.70
Os02g05470*	2	3 695	483	51.09	9.16	At1g25440	1	1 734	417	46.93	4.99
Os02g08150	2	1 687	350	35.97	6.66	At1g28050	1	2 283	433	47.61	5.98
Os02g39710	2	1 544	333	34.95	4.93	At1g49130*	1	1 224	326	38.32	6.43
Os02g40510	2	3 267	519	57.88	6.04	At1g51600*	1	3 321	302	33.13	7.19
Os02g49230*	2	5 416	481	52.43	6.09	At1g63820	1	1 946	293	33.62	4.61
Os02g49880	2	2 986	453	49.24	6.26	At1g68520	1	1 794	406	46.06	5.49
Os03g04620*	3	4 030	412	45.21	4.17	At1g73870	1	1 499	392	44.02	4.66
Os03g17570*	3	9 212	768	84.08	6.39	At2g24790	2	1 248	294	32.32	6.65
Os03g22770*	3	2 585	404	43.44	5.30	At2g32310	2	1 706	325	36.91	4.80
Os03g47970*	3	4 413	320	34.12	4.29	At2g33350*	2	3 213	410	45.48	4.39
Os03g50310	3	2 759	422	45.80	6.66	At2g33500*	2	2 268	402	44.52	7.24
Os03g52450*	3	4 801	272	29.03	6.51	At2g46670	2	795	183	20.54	10.48
Os04g42020	4	1 505	334	34.60	4.86	At2g46790*	2	2 315	468	52.57	7.77
Os05g38990	5	1 576	325	34.94	5.80	At2g47890*	2	1 637	332	37.51	5.97
Os05g51690*	5	3 772	309	34.17	4.51	At3g02380	3	1 451	347	38.52	5.64
Os06g01340	6	904	224	24.01	8.83	At3g07650*	3	2 860	372	40.75	6.23
Os06g15330	6	2 119	449	48.09	4.88	At3g12890	3	1 271	251	28.86	5.74
Os06g16370	6	2 285	396	43.08	4.59	At3g21175*	3	2 870	297	32.88	6.80
Os06g19444*	6	6 614	409	43.94	4.88	At3g21880	3	1 346	364	40.47	5.55
Os06g44450	6	1 690	372	39.19	6.62	At4g15250	4	1 165	330	36.62	4.98
Os06g48534	6	4 386	293	31.57	4.69	At4g24470*	4	2 320	309	33.26	5.99
Os06g48610	6	4 295	467	49.31	7.12	At4g25990*	4	1 607	394	44.58	8.57
Os07g15770	7	2 784	288	30.65	6.51	At4g27900*	4	2 286	261	29.45	4.60
Os07g47140	7	3 050	381	40.36	4.83	At5g02810	5	4 081	727	79.66	7.80
Os07g49460*	7	12 520	743	79.92	8.86	At5g14370	5	1 823	339	38.28	9.89
Os08g15050	8	1 308	304	29.94	6.50	At5g15840*	5	1 577	373	41.99	6.96
Os08g42440	8	4 358	489	51.86	6.63	At5g15850	5	1 807	355	39.49	6.50
Os09g06464	9	16 550	336	36.08	5.27	At5g24470	5	2 892	667	74.84	9.21
Os09g33550	9	3 254	472	50.59	6.77	At5g24930	5	2 003	406	44.68	7.19
Os09g36220*	9	4 672	624	69.62	7.07	At5g41380	5	2 455	307	35.08	5.08
Os10g32900*	10	2 936	450	47.05	4.23	At5g48250	5	2 404	373	41.25	5.97
Os10g41100	10	2 821	298	30.21	7.85	At5g53420*	5	2 204	264	30.04	4.57
Os11g01074*	11	8 681	717	61.80	4.52	At5g57180*	5	2 982	435	48.41	7.56
Os11g01100	11	935	154	16.35	7.79	At5g57660	5	1 526	355	38.82	6.43
Os11g05930*	11	4 740	700	76.09	8.01	At5g59990	5	1 879	241	26.59	6.51
Os12g01080	12	533	147	16.19	8.70	At5g60100	5	2 288	495	55.29	7.62
Os12g01100	12	935	154	16.37	7.79	At5g61380	5	3 464	618	69.20	7.59
Os12g16160	12	1 188	221	24.03	7.50	Sb01g029080	1	2 728	298	30.21	7.24
Zm2g004483	9	2 891	275	28.41	6.41	Sb04g005250	4	1 754	376	39.06	6.40
Zm2g021777	5	1 544	335	35.39	5.04	Sb04g025660	4	1 587	336	35.35	5.00
Zm2g038783	9	1 820	364	38.81	5.81	Sb04g029180	4	1 504	289	33.02	8.24
Zm2g092363	10	1 406	278	28.16	6.08	Sb06g000570	6	2 643	245	26.78	6.21
Zm2g095598	2	1 507	320	34.07	5.43	Sb06g021480	6	1 296	329	34.52	5.19
Zm2g381691	10	2 555	240	26.21	5.84	Sb07g008550	7	1 378	292	29.11	5.84
AC2333888	10	1 103	323	34.01	5.20	Sb10g010050	10	1 490	411	44.25	5.38
						Sb10g026060	10	1 801	376	40.03	5.97

1)“Os”代表水稻 CCT 结构域基因，“Zm”代表玉米 CCT 结构域基因，“At”代表拟南芥 CCT 结构域基因，“Sb”代表高粱 CCT 结构域基因；基因编号右上方端带“*”标示的表示含不止一种转录模式。“Os”，“Zm”，“At”，“Sb” represents CCT domain genes of rice, *Arabidopsis*, sorghum, maize. “*” , more than one transcript pattern.

为了更好地研究含 CCT 结构域基因的进化, 结果表明: 在水稻和拟南芥基因组中 CCT 结构域基
析了水稻和拟南芥的染色体片断之间的复制事件。因在染色体上都存在串联重复和大片断的复制。水

稻中,在第 11 和 12 染色体上存在 CCT 结构域基因串联成簇分布的情况,并且在不同染色体之间存在 10 次复制事件,如第 1 染色体的 Os01g61900 与第 5 染色体的 Os05g38990,第 3 染色体的 Os03g22770 与第 7 染色体的 Os07g47140 都可能存在着复制关系。位于第 2 染色体上串联重复的 2 个基因(Os02g05470

和 Os02g05510)和第 6 染色体串联重复的 2 个基因(Os06g48610 和 Os06g48534)存在染色体复制关系(图 1)。而在拟南芥中,在第 2 和第 5 染色体上皆存在串联分布的 CCT 结构域基因簇,在不同染色体之间发生了 8 次复制事件,而在第 1 染色体长臂和短臂之间也发生过 1 次复制(图 2)。

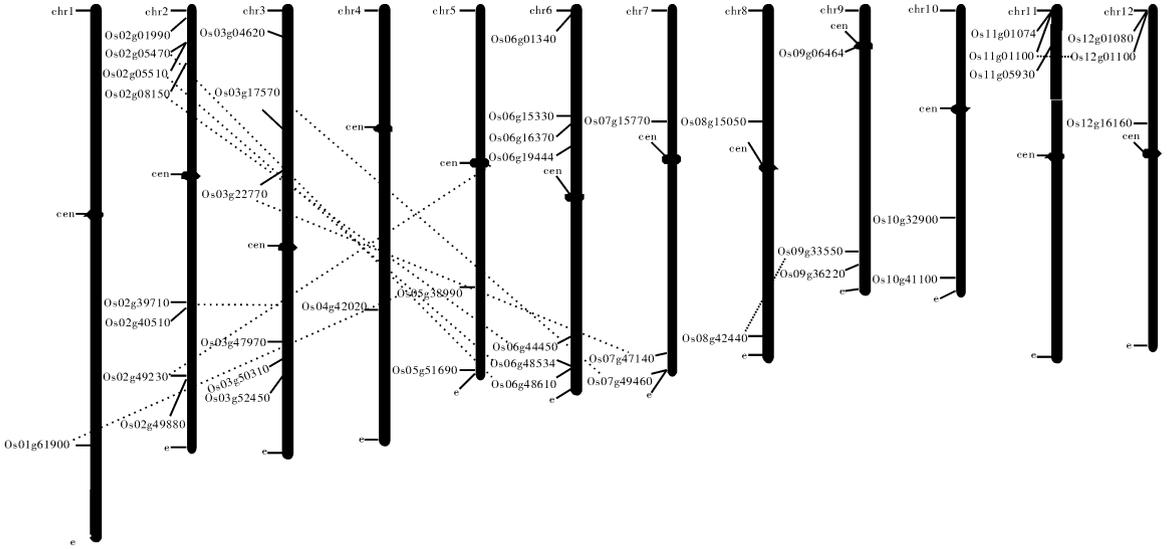


图 1 水稻 CCT 结构域基因染色体定位及片段复制分布图(染色体片段复制关系用虚线表示)

Fig. 1 Localization of 41 CCT genes on the rice chromosomes (Genes lying on duplicated segments of genome have been joined by dashed lines)

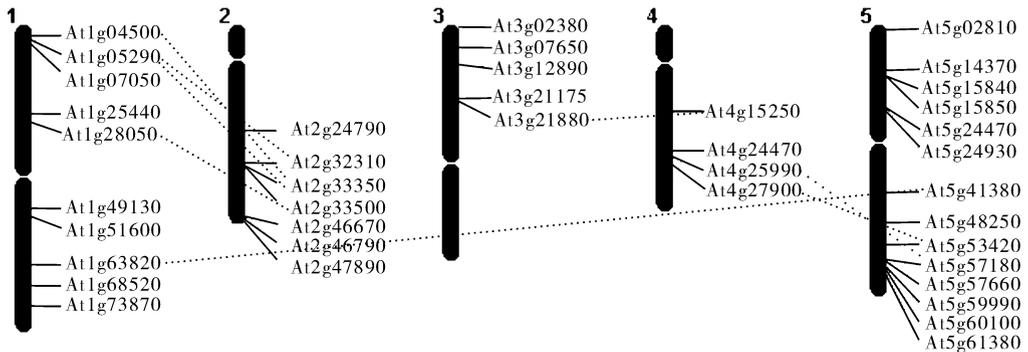


图 2 拟南芥 CCT 结构域基因染色体定位及片段复制分布图(染色体片段复制关系用虚线表示)

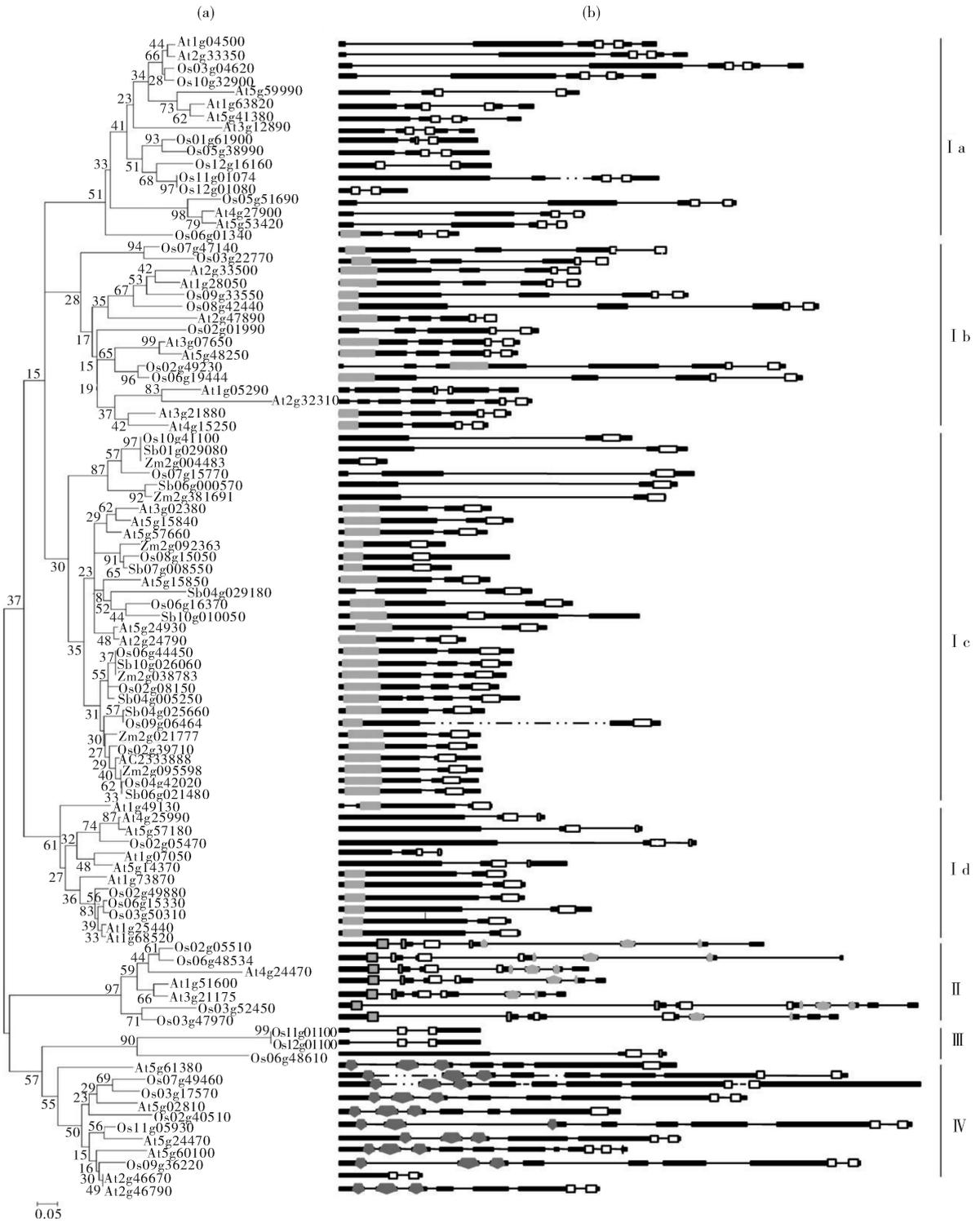
Fig. 2 Localization of 40 CCT genes on the *Arabidopsis* chromosomes (Duplicated segments of genome have been joined by dashed lines)

2.3 CCT 结构域基因结构和蛋白质分析

根据 TIGR 对水稻的基因组注释和 PLANT-TFDB 数据库对拟南芥 CCT 结构域基因的注释,分析了 CCT 结构域基因的基因结构。结果表明:在 4 个物种中,来自水稻、玉米和高粱各有 1 个基因不含有内含子,其余基因都含有 1~9 个内含子。

CCT 结构域基因编码蛋白质的长度和分子质量分布范围很广,最短的蛋白质是水稻基因

Os12g01080 编码的,仅仅有 147 个氨基酸,而最长的则为水稻基因 Os03g17570 编码的,含有多达 768 个氨基酸;蛋白质的分子质量也从 16.19 ku 到 84.08 ku 不等,并且水稻的 CCT 蛋白质大小差异比拟南芥的更为显著。蛋白质等电点分布同样十分广泛,涵盖了从酸性到碱性的蛋白质。等电点最低的 3 个蛋白质都由水稻编码,而最高的 3 个蛋白质则全为拟南芥编码,拟南芥的 CCT 结构域基因编码



a: CCT 保守结构域序列邻接法构建的系统进化树 Phylogenetic analysis of the coserved CCT domain;

b: CCT 结构域基因结构及保守结构域图示 Structures and other coserved domians of CCT gene proteins.

■:外显子 Exon; - - -:内含子 Intron; - - - -:按 5 倍比例缩放的内含子 Large intron; □: CCT 保守结构域 CCT domain; ▨: B-box 保守结构域 B-box domain; ▩: Tify 保守结构域 Tify domain; ◆: GATA 保守结构域 GATA domain; ◆: REC 保守结构域 REC domain.

图 3 拟南芥、水稻、玉米、高粱 CCT 结构域的系统进化分析(a)及基因结构图示(b)

Fig. 3 Phylogenetic analysis and structures of CCT domian genes of *Arabidopsis*, rice, sorghum, maize

的蛋白质具有更为广泛的等电点变异范围。

通过对 CCT 结构域基因编码蛋白质的结构域分析,表明除了都编码 CCT 结构域外,还包含其他 4 种结构域: B-box, REC, GATA 以及 TIFY(图 3)。最早克隆的拟南芥的 CO 蛋白质存在着 CCT 结构域和 B-box 结构域,而水稻的 6 个蛋白质,拟南芥的 10 个蛋白质,高粱的 5 个蛋白质和玉米的 4 个蛋白质的 N 端都鉴定到了 2 个 B-box 结构域,此外水稻和拟南芥中还有 18 个蛋白质仅仅包含 1 个 B-box 结构域。与 CO 蛋白质不同,在 4 个物种中共有多达 52 个 CCT 蛋白质不编码 B-box 结构域。此外, REC 结构域存在于水稻和拟南芥的 10 个蛋白质中,而在水稻和拟南芥中还发现分别有 4 个和 3 个蛋白质编码有 GATA 和 TIFY 结构域,这 2 种结构域在玉米和高粱的 CCT 蛋白质中并不存在。

上述结果表明, CCT 结构域基因家族不仅基因结构变异较大,而且其编码的蛋白质有较大的极性变化,并含有丰富的结构域类型。

2.4 CCT 结构域基因系统进化树的构建

为了研究 CCT 结构域基因在各物种间的进化关系,将 CCT 结构域的氨基酸序列利用软件 MEGA 4.0 中 N-J 的方法构建系统进化树,结果是 4 个物种的 CCT 结构域基因可以根据序列的相似程度分为 4 组: I、II、III、IV。组 I 可被分为 4 个亚组: Ia、Ib、Ic、Id(图 3),组 III 中仅包含有来自水稻的基因,而来自玉米和高粱的基因则全部集中在亚组 Ic 中,其余的组别则全部由水稻和拟南芥的部分基因共同组成,其中组 IV 中来自拟南芥的基因都定位于第 2 和第 5 染色体上。亚组 Ic 中包含 4 个物种的基因,有 6 个来自拟南芥的基因在 Ic 亚组中出现,在禾本科没有同源性较高基因存在;而 3 个禾本科物种之间则有亲缘关系较近的基因对存在,比如水稻 Os10g41100 与玉米 Zm2g004483、高粱 Sb01g029080 亲缘关系最近,玉米与高粱还含有 2 个同源性较高的配对基因;由此可见,该亚组类蛋白质中,出现了单双子叶植物之间的序列分化,鉴于 CO 和 Hd1 作为直系同源基因,但却在拟南芥和水稻之间出现了功能分化,同时水稻另外 1 个控制抽穗期的 QTL 基因 *Ghd7* 也分布在此亚组中,因而该亚组类基因可能对拟南芥和禾本科植物之间开花机制的分歧进化起着重要的作用。

利用含 CCT 结构域基因的蛋白质的全长序列

进行系统进化树构建可以得到与上面基本相同的系统树,只有来自拟南芥的基因 At2g32310 分类有所不同,但是仍然与玉米、高粱、水稻无同源关系较近基因,因此 CCT 结构域的相似性可以准确地反应 CCT 结构域基因之间的进化关系。

3 讨论

在双子叶长日照植物拟南芥中,光周期通过生物钟相关基因 *GI* 调控 *CO* 的节律性表达^[17], *CO* 通过调控 *FT* 和 *SOC1* 的表达而最终控制成花转换^[18-20],在单子叶短日照植物水稻中也存在类似的 *OsGI/Hd1/Hd3* 途径^[3]。大量的研究表明:虽然 CCT 结构域基因广泛地参与对开花时间的控制,但是对日长的响应和对开花起促进或抑制作用的方面却发生了种内和种间的分化。

多基因家族成员在进化过程往往会发生功能分化,获得或者失去特定的保守结构域是其中一个重要的原因。拟南芥 CO 蛋白质仅包含有 B-box 和 CCT 2 种结构域,但是在本研究中,笔者在 4 个物种的 CCT 结构域基因编码的蛋白质中共发现了 5 种功能结构域。B-box 结构域属于锌指结构蛋白质的一种,可能参与蛋白质-蛋白质相互作用。REC 结构域功能目前尚不明确。GATA 结构域编码锌指结构蛋白质,具有结合 GATA 模体(motif)的功能^[21],这个结构域的存在,必然导致包含有此结构域的 CCT 蛋白如 Os02g05510、Os03g47970、At1g51600 等相对于其他成员具有更为广泛的下游靶基因,从而具有更为广泛的调控功能。此外,4 个物种有多达 52 个成员仅仅具有 CCT 结构域,这些蛋白质与 CO 蛋白质相比,缺少了 B-box 结构域,也可能导致这些蛋白质与 CO 蛋白质具有不同的调控功能。因此不同的结构域的获得与丢失为 CCT 结构域基因的功能分化提供了变异的基础。

另一方面,不同物种可以通过染色体的倍增和基因的复制实现基因数目的增加,最终形成多基因家族。新形成的成员可以通过积累变异从而与原始的同源基因发生功能分化。本研究中 CCT 结构域基因共被分为 4 组,其中组 III 中仅包含有来自水稻的基因,这表明水稻的基因组已经进化出了物种特异的 CCT 结构域基因。在亚组 Ic 中虽然包含有 4 个物种的基因,但是水稻的 *Hd1*(Os06g16370)和拟南芥中的直系同源基因 *CO*(At5g15840)以及

水稻的 *Ghd7* (Os07g15770) 和 *Hd1* 之间都已经发生了功能的分化。亲缘关系如此之近而功能尚且不同,因此笔者推测,组 III 中水稻特异的 CCT 结构域基因可能具有物种特异的功能。Ic 亚组虽然包含了部分的拟南芥基因,但是其基本上处于一个独特的亚组分枝上,其余的全是禾本科基因,包含了全部的玉米和高粱的基因,鉴于 *CO* 和水稻 *Hd1* 功能的分化,此组中的 CCT 结构域基因可能对禾本科作物开花时间的控制方面具有特异的功能。水稻中短日照抑制开花的基因 *OsCO3* (Os09g06464)、长短日照均抑制开花的基因 *OsCOL4* (Os02g39710) 位于亚组 Ic 中一个分支,推测该分支类基因开始出现长日照抑制水稻开花期的功能分化。因此基因的复制和染色体的倍增对 CCT 结构域基因在物种间和物种内的功能分化也有相当的贡献。

相对于蛋白质结构的变化,基因表达模式的变异在物种进化中同样发挥重要的作用,是同源基因功能分化的一个非常重要的因素^[22]。CCT 结构域基因在进化过程中同样发生了表达模式的变异。例如水稻的 *Ghd7* 在长日照条件下比短日照条件具有更高的表达量^[11],而同样分布在 Ic 亚组中的 *Hd1* 的表达模式则不随日照长短的变化而改变^[10],因此二者对日长具有不同的响应,这表明这 2 个基因的启动子具有不同的变异。推测基于启动子的变异,不同的 CCT 结构域基因成员对光周期、节律性和日照长短可能具有不同的响应,从而促进了 CCT 结构域基因家族的功能分化。

CCT 结构域基因成员基因表达模式的变异和蛋白质结构的改变共同导致了 CCT 结构域基因在物种间和物种内的功能分化。这种改变使得不同物种或者同一物种的不同种群能够适应它们的地理分布特点,保证不同物种或者种群正确地响应调控开花的环境因子,自主选择合适的开花时间,从而成功地产生种子,繁衍后代。

参 考 文 献

- [1] SHINYOUNG L, GYNHEUNG A. Diversified mechanisms for regulating flowering time in a short-day plant rice[J]. *Journal of Plant Biology*, 2007, 50(3): 241-248.
- [2] HIROYUKI T, SHOJIRO T, REINA K, et al. Florigen and the photoperiodic control of flowering in rice[J]. *Rice*, 2008(1): 25-35.
- [3] KOMIYA R, YOKOI S, SHIMAMOTO K. A gene network for long-day flowering activates RFT1 encoding a mobile flowering signal in rice[J]. *Development*, 2009, 136(20): 3443-3450.
- [4] TAKATO I. Arabidopsis circadian clock and photoperiodism: time to think about location[J]. *Curr Opin Plant Biol*, 2010, 13(1): 83-89.
- [5] KOORNNEEF M, ALONSO-BLANCO C, BLANKESTIJN-DEVRIES H, et al. Genetic interactions among late-flowering mutants of Arabidopsis[J]. *Genetics*, 1998, 148(2): 885-892.
- [6] PUTTERILL J, LAURIE R, MACKNIGHT R. It's time to flower: the genetic control of flowering time [J]. *Bioessays*, 2004, 26(4): 363-373.
- [7] PUTTERILL J, ROBSON F, LEE K, et al. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors [J]. *Cell*, 1995, 80(6): 847-857.
- [8] STRAYER C, OYAMA T, SCHULTZ T F, et al. Cloning of the *Arabidopsis* clock gene *TOC1*, an autoregulatory response regulator homolog[J]. *Science*, 2000, 289(5480): 768-771.
- [9] ROBSON F, COSTA M M, HEPWORTH S R, et al. Functional importance of conserved domains in the flowering-time gene *CONSTANS* demonstrated by analysis of mutant alleles and transgenic plants[J]. *Plant J*, 2001, 28(6): 619-631.
- [10] WENKEL S, TURCK F, SINGER K, et al. *CONSTANS* and the CCAAT Box Binding Complex share a functionally important domain and interact to regulate flowering of *Arabidopsis* [J]. *Plant Cell*, 2006, 18(11): 2971-2984.
- [11] YANO M, KOJIMA S, TAKAHASHI Y, et al. Genetic control of flowering time in rice, a short-day plant[J]. *Plant Physiology*, 2001, 127(4): 1425-1429.
- [12] XUE W, XING Y, ZHANG Q, et al. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice[J]. *Nat Genet*, 2008, 40(6): 761-767.
- [13] KIM S K, YUN C H, LEE J H, et al. *OsCO3*, a *CONSTANS-LIKE* gene, controls flowering by negatively regulating the expression of *FT*-like genes under SD conditions in rice[J]. *Planta*, 2008, 228(2): 355-365.
- [14] LEE Y S, JEONG D H, LEE D Y, et al. *OsCOL4* is a constitutive flowering repressor upstream of *Ehd1* and downstream of *OsphyB*[J]. *Plant J*, 2010, 63(1): 18-30.
- [15] GRIFFITHS S, DUNFORD R P, COUPLAND G, et al. The evolution of *CONSTANS*-like gene families in barley, rice, and *Arabidopsis*[J]. *Plant Physiology*, 2003, 131(4): 1855-1867.
- [16] TAMURA K, DUDLEY J, NEI M, et al. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0 [J]. *Mol Biol Evol*, 2007, 24(8): 1596-1599.
- [17] SAWA M, NUSINOW D A, IMAIZUMI T, et al. FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis* [J]. *Science*, 2007, 318(5848):

- 261-265.
- [18] SIMON R, IGENO M I, COUPLAND G. Activation of floral meristem identity genes in *Arabidopsis*[J]. *Nature*, 1996, 384 (6604): 59-62.
- [19] KARDAILSKY I, SHUKLA V K, AHN J H. Activation tagging of the floral inducer *FT*[J]. *Science*, 1999, 286 (5446): 1962-1965.
- [20] SAMACH A, ONOUCHI H, GOLD S E, et al. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*[J]. *Science*, 2000, 288(5471): 1613-1616.
- [21] OMICHINSKI J G, CLORE G M, SCHAAD O. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1[J]. *Science*, 1993, 261(5120): 438-446.
- [22] DOEBLEY J F, GAUT B S, SMITH B D. The molecular genetics of crop domestication [J]. *Cell*, 2006, 127 (7): 1309-1321.

Sequence Evolution Analysis of CCT Domain Gene Family in Rice, *Arabidopsis*, Maize and Sorghum

CHEN Hua-xia SHEN Guo-jing WANG Lei XING Yong-zhong

College of Life Science and Technology/National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

Abstract The CCT domain defines a conserved 43~45 amino acids of the 3' region of CONSTANS, CO-like, and TOC1. The plant specific CCT domain gene plays a role in light signaling and regulation of photoperiodic flowering. Meanwhile, the evolutionary relationship among family members has not been well studied. In this study, 97 CCT domain containing genes were retrieved from four completely sequenced and annotated genomes. Analyses of corresponding protein structure and genetic relationships showed that CCT domain containing protein genes kept a wide range of variation within and among different species; Sizes of CCT domain gene family were increased through chromosome duplication. CCT domain gene family was classified into four subgroups according to the CCT domain. One sub-group gathered most of the grass CCT domain gene, which might be involved in the regulation of flowering time. The functional differentiation within and among species was comprehensively caused by the expansion of gene family members, structural variation of protein and patterns of gene expression.

Key words CCT domain gene family; chromosome duplication; protein structure; function differentiation

(责任编辑:陆文昌)