

拟南芥和水稻 *CesA* 基因家族的生物信息学分析

王振怡,王金朋,潘玉欣,张家琦,王希胤*
(河北联合大学 生命科学学院 基因组学与计算生物学研究中心,河北 唐山 063009)

摘要:以拟南芥和水稻 21 个纤维素合成酶基因(*CesA*)为研究对象,对其基因结构、保守结构域、蛋白质跨膜结构、亚细胞定位、基因表达等进行综合分析。结果显示,*CesA* 基因长度在 3.5 ~ 6.5 kb,有 7 ~ 13 个内含子(*OsCesA10* 除外)。*CesA* 基因编码的蛋白质保守性较强,含有该基因家族的保守结构域 Ring finger(*OsCesA10*、*OsCesA11* 除外)和 Cellulose synthase。水稻和拟南芥的 *CesA* 蛋白跨膜螺旋数量为 6 个或 8 个(*OsCesA10* 除外)。蛋白质亚细胞定位结果表明,21 条 *CesA* 蛋白全部定位在质膜上。大部分 *AtCesA* 基因在植株的幼根、幼叶和愈伤组织中均有表达,而大部分 *OsCesA* 基因仅在幼根或幼叶中表达。

关键词:纤维素合成酶基因;拟南芥;水稻;生物信息学;基因表达

中图分类号: Q51;S511 **文献标志码:** A **文章编号:** 1004 - 3268(2015)06 - 0013 - 05

Bioinformatic Analysis of *CesA* Gene Family in *Arabidopsis* and Rice
WANG Zhenyi, WANG Jinpeng, PAN Yuxin, ZHANG Jiaqi, WANG Xiyin*
(Center for Genomics and Computational Biology, College of Life Sciences, Hebei United University, Tangshan 063009, China)

Abstract: In this study, twenty-one cellulose synthase genes(*CesA*) from *Arabidopsis* and rice were analyzed in their structures, conservative motifs, protein transmembrane helices, protein subcellular locations and gene expressions. The results showed that the size of *CesA* ranged from 3.5 to 6.5 kb, with 7—13 small introns(except *OsCesA10*). *CesA* had highly conserved structure with the family domain “Ring finger”(except *OsCesA10* and *OsCesA11*) and “Cellulose synthase”. The number of protein transmembrane helices in *Arabidopsis* and rice was 6 or 8(except *OsCesA10*), respectively. The subcellular locations showed that the 21 *CesA* proteins were all located on the membranes. Most of the *AtCesA* genes expressed in young root, young leaves and calluses, whereas most of the *OsCesA* expressed only in young root or young leaves.

Key words: cellulose synthase gene(*CesA*); *Arabidopsis*; rice; bioinformatics; gene expression

纤维素是植物细胞壁的主要组成部分,其决定了细胞的扩展方向,并为植物细胞提供支持材料^[1]。自然界每年有 1 800 亿 t 的纤维素生成,是地球上最丰富的生物大分子和重要的可再生资源^[2]。研究纤维素合成机制对扩大生物质材料和生物质能源资源意义重大。

许多研究表明,植物纤维素的生物合成是由植物纤维素合成酶(*CesA*)组成的纤维素合成酶复合体(Rosette)和其他酶共同完成的一个复杂过程^[2-3]。Pear 等^[4]运用同源克隆的方法首次从棉纤维中克隆了 *GhCesA1* 和 *GhCesA2*。目前研究者们已从 40 多种植物中克隆了 1 400 多个相关的序列并

收稿日期:2014 - 12 - 16
基金项目:国家自然科学基金项目(31170212);河北省教育厅项目(Y2012025);河北省科学技术厅项目(14962905D);河北联合大学自然科学研究基金项目(z201235,z201230)
作者简介:王振怡(1983 -),女,河北正定人,讲师,硕士,主要从事生物信息学、比较基因组学研究。
E - mail:zhenyiwang0301@163.com
* 通讯作者:王希胤(1972 -),男,河北唐山人,教授,博士,主要从事生物信息学、比较基因组学研究。
E - mail:wang.xiyin@gmail.com

发布在公共数据库中^[5-6],如拟南芥(*Arabidopsis thaliana*)^[3]、玉米(maize)^[7]、大麦(barley)^[8]、杨树(poplar)^[9-10]、松(pine)^[11]、水稻(rice)^[12-13]、细叶桉(*Eucalyptus tereticornis*)^[14]、紫花苜蓿(Alfalfa)^[15]等。植物纤维素合成酶基因家族 *CesA* 是一个庞大的基因家族,采用生物信息学方法对拟南芥和水稻中的 *CesA* 基因家族成员进行基因结构、保守结构域、蛋白质跨膜结构、亚细胞定位、表达模式等分析,旨在揭示植物 *CesA* 基因家族的结构特点和进化历史,为后续研究 *CesA* 基因家族的多样性提供理论基础。

1 材料和方法

1.1 材料

拟南芥和水稻的 *CesA* 基因核酸以及对应的蛋白质序列数据分别从拟南芥数据库 TAIR(<http://www.arabidopsis.org/>)和水稻数据库 Rice Genome Annotation Project(<http://rice.plantbiology.msu.edu/index.shtml>)中搜索下载得到。

1.2 方法

1.2.1 *CesA* 基因序列和结构分析 利用在线基因结构显示系统 GSDS(gene structure display server)(<http://gsds.cbi.pku.edu.cn/>)对拟南芥和水稻 *CesA* 基因进行结构预测^[16]。

1.2.2 *CesA* 蛋白的保守结构域分析 利用在线分

析系统 SMART analysis service(<http://smart.embl-heidelberg.de/>)对水稻和拟南芥的 *CesA* 蛋白结构域进行分析。

1.2.3 *CesA* 蛋白的跨膜结构和亚细胞定位预测 通过在线 TMHMM Server V. 2. 0(<http://www.cbs.dtu.dk/services/TMHMM>)分析系统^[17-18]预测水稻和拟南芥 *CesA* 蛋白的跨膜结构,蛋白质的亚细胞定位预测通过在线 WoLF PSORT 系统^[19](<http://wolfpsort.org/>)完成。

1.2.4 *CesA* 基因表达模式的分析 拟南芥和水稻 *CesA* 基因家族表达模式预测在网站 MPSS program(<http://mpss.fscj.edu/>)上进行,分析幼根、幼叶和愈伤组织中的表达量差异。

2 结果与分析

2.1 水稻和拟南芥 *CesA* 基因的结构分析

从拟南芥和水稻基因组数据库中共搜索到 21 个 *CesA* 基因,其中拟南芥中有 10 个 *CesA* 基因(*At-CesA1—AtCesA10*),3 个分布在第 4 号染色体上,5 个分布在第 5 号染色体上,2 个分布在第 2 号染色体上;水稻中有 11 个 *CesA* 基因(*OsCesA1—OsCesA11*),3 个分布在第 7 号染色体上,2 个分布在第 3 号染色体上,其他 6 个各分布在不同的水稻染色体上(表 1)。

表 1 *AtCesA* 和 *OsCesA* 基因家族

编号	基因名称	基因位点	蛋白质特性		
			蛋白质跨膜螺旋数	亚细胞定位	保守结构域
1	<i>AtCesA1</i>	AT4G32410	6	Plasma membrane	Ring finger、Cellulose synthase
2	<i>AtCesA2</i>	AT4G39350	6	Plasma membrane	Ring finger、Cellulose synthase
3	<i>AtCesA3</i>	AT5G05170	8	Plasma membrane	Ring finger、Cellulose synthase
4	<i>AtCesA4</i>	AT5G44030	8	Plasma membrane	Ring finger、Cellulose synthase
5	<i>AtCesA5</i>	AT5G09870	8	Plasma membrane	Ring finger、Cellulose synthase
6	<i>AtCesA6</i>	AT5G64740	8	Plasma membrane	Ring finger、Cellulose synthase
7	<i>AtCesA7</i>	AT5G17420	6	Plasma membrane	Ring finger、Cellulose synthase
8	<i>AtCesA8</i>	AT4G18780	8	Plasma membrane	Ring finger、Cellulose synthase
9	<i>AtCesA9</i>	AT2G21770	6	Plasma membrane	Ring finger、Cellulose synthase
10	<i>AtCesA10</i>	AT2G25540	8	Plasma membrane	Ring finger、Cellulose synthase
11	<i>OsCesA1</i>	LOC_Os05g08370	8	Plasma membrane	Ring finger、Cellulose synthase
12	<i>OsCesA2</i>	LOC_Os03g59340	6	Plasma membrane	Ring finger、Cellulose synthase
13	<i>OsCesA3</i>	LOC_Os07g24190	8	Plasma membrane	Ring finger、Cellulose synthase
14	<i>OsCesA4</i>	LOC_Os01g54620	8	Plasma membrane	Ring finger、Cellulose synthase
15	<i>OsCesA5</i>	LOC_Os03g62090	8	Plasma membrane	Ring finger、Cellulose synthase
16	<i>OsCesA6</i>	LOC_Os07g14850	8	Plasma membrane	Ring finger、Cellulose synthase
17	<i>OsCesA7</i>	LOC_Os10g32980	6	Plasma membrane	Ring finger、Cellulose synthase
18	<i>OsCesA8</i>	LOC_Os07g10770	6	Plasma membrane	Ring finger、Cellulose synthase
19	<i>OsCesA9</i>	LOC_Os09g25490	6	Plasma membrane	Ring finger、Cellulose synthase
20	<i>OsCesA10</i>	LOC_Os12g29300	0	Plasma membrane	Cellulose synthase
21	<i>OsCesA11</i>	LOC_Os06g39970	6	Plasma membrane	Cellulose synthase

比较 *CesA* 基因的基因组序列和编码序列,结果表明,21 个 *CesA* 基因的长度在 3.5 ~ 6.5 kb,内含子数量为 7 ~ 13 个(*OsCesA10* 除外)。大部分 *CesA* 基因含有 11、12、13 个内含子(81%)。拟南芥 *AtCesA4*、*AtCesA7*、*AtCesA8* 含有 11 个内含子,*AtCesA2*、*AtCesA5*、*AtCesA6*、*AtCesA9* 中含有 12 个内含子,*AtCesA1*、*AtCesA3*、*AtCesA10* 含有 13 个内含子;水稻 11 个 *CesA* 基因中,*OsCesA7* 和 *OsCesA9* 的内含子数目分别为 8 个和 10 个,*OsCesA10* 和 *OsCesA11* 的内含子分别为 1 个和

7 个,其他 7 个基因基因的内含子数目为 12、13 个(图 1)。*CesA* 基因转录产物的长度是 2.5 ~ 3.5 kb,编码的肽链长度约为 861 ~ 1 097 个氨基酸(*OsCesA10* 除外)。*OsCesA10* 和 *OsCesA11* 两个基因还未测得基因的全长序列,其蛋白质的氨基酸数目和基因的外显子数目远小于其他的 *OsCesA* 基因。*CesA* 基因内含子和外显子的边界区域是高度保守的,基因结构的差异主要取决于内含子的多少。

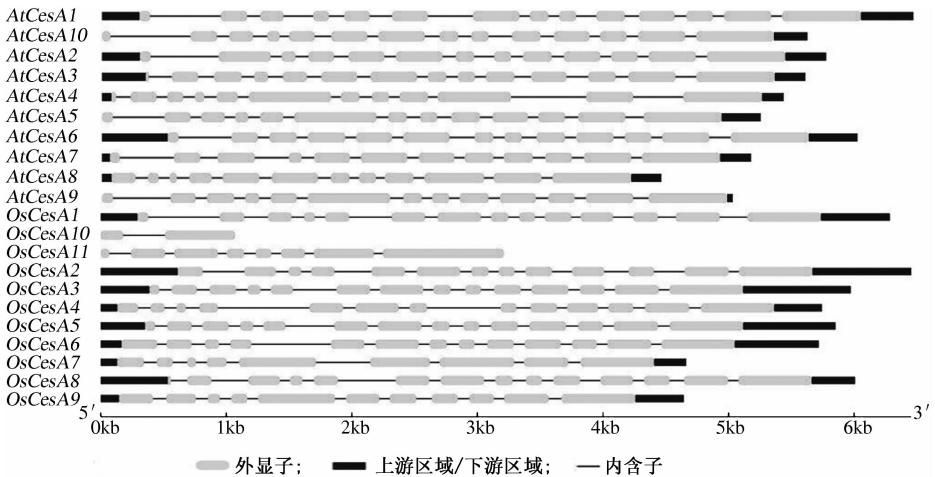


图 1 拟南芥和水稻 *CesA* 基因结构分布

2.2 水稻和拟南芥中 *CesA* 蛋白的结构域分析

利用 SMART 在线分析系统对水稻和拟南芥的 21 条 *CesA* 蛋白序列进行结构功能域分析,结果显示,有 19 条 *CesA* 蛋白序列包括该家族的保守结构域 Ring finger 和 Cellulose synthase,只有 *OsCesA10* 和 *OsCesA11* 没有 Ring finger(表 1)。Ring finger 是 40 ~ 60 残基绑定了 2 个锌原子的一种特殊类型,并且其可能参与介导蛋白质 - 蛋白质之间的关系。

Ring finger 存在于氨基末端,一些研究显示它在蛋白质 - 蛋白质相互作用中发挥了重要作用^[20]。

2.3 水稻和拟南芥中 *CesA* 蛋白的跨膜结构和亚细胞定位预测

根据蛋白质跨膜预测系统 TMHMM Server V. 2. 0 分析,水稻和拟南芥的 21 条 *CesA* 蛋白跨膜螺旋数量为 6 个或 8 个(*OsCesA10* 除外,表 1)。从图 2 可以看出,*AtCesA1* 蛋白的跨膜螺旋的数量是 6 个,*OsCesA1*

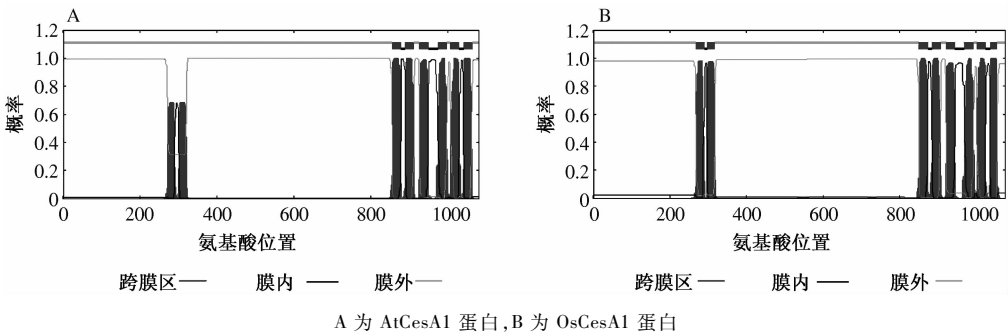


图 2 拟南芥 *AtCesA1* 和水稻 *OsCesA1* 蛋白跨膜区预测结果

跨膜螺旋的数量是 8 个。蛋白质亚细胞定位预测结果表明,21 条 *CesA* 蛋白全部定位在质膜上(表 1)。有研究认为,纤维素多糖的合成是在质膜上进行的,而非纤维素多糖的合成则是在高尔基体上进行

的^[12]。*CesA* 蛋白作为蛋白质复合体的一个成员,其活动是在质膜表面进行的,通过电子显微镜可以观察到,观察到的这种结构被称为“玫瑰花状结构”。

2.4 Cesa 基因在幼根、幼叶和愈伤组织中的表达分析

植物体内存在多种参与纤维素合成的纤维素合成酶,不同纤维素合成酶参与不同组织或不同细胞壁结构中的纤维素合成,具有组织表达的特异性^[21-25]。从 MPSS 网站中,选择幼根,幼叶和愈伤组织进行组织特异性表达分析,结果(表 2)表明,大部分 *AtCesA* 基因在幼根、幼叶和愈伤组织中均有表达。*AtCesA8* 基因在幼根和幼叶中表达,*AtCesA9* 基因只在幼叶中表达,*AtCesA10* 基因在这 3 个组织中均无表达,拟南芥中其余 7 个 *CesA* 基因(*AtCesA1*—*AtCesA7*)在这 3 个组织中均有表达。而大部分 *OsCesA* 在幼根或幼叶中表达,*OsCesA2*、*OsCesA6* 和 *OsCesA8* 只在幼根中表达,*OsCesA3* 只在幼叶中表达,只有 *OsCesA1* 在幼根、幼叶和愈伤组织中均有表达,*OsCesA10* 和 *OsCesA11* 在这 3 个组织中均无表达,其他 4 个 *OsCesA* 在其中的 2 个组织中均有表达。

表 2 拟南芥和水稻 CesA 蛋白表达分析

蛋白质名称	幼根	幼叶	愈伤组织
<i>AtCesA1</i>	1	3	1
<i>AtCesA2</i>	1	3	1
<i>AtCesA3</i>	3	1	1
<i>AtCesA4</i>	3	1	1
<i>AtCesA5</i>	1	3	1
<i>AtCesA6</i>	3	1	1
<i>AtCesA7</i>	3	1	1
<i>AtCesA8</i>	3	1	0
<i>AtCesA9</i>	0	3	0
<i>AtCesA10</i>	0	0	0
<i>OsCesA1</i>	1	1	3
<i>OsCesA2</i>	3	0	0
<i>OsCesA3</i>	0	3	0
<i>OsCesA4</i>	3	1	0
<i>OsCesA5</i>	3	0	1
<i>OsCesA6</i>	3	0	0
<i>OsCesA7</i>	3	1	0
<i>OsCesA8</i>	3	0	0
<i>OsCesA9</i>	3	1	0
<i>OsCesA10</i>	0	0	0
<i>OsCesA11</i>	0	0	0

注:0 表示没有表达,1 表示表达,3 表示表达量最高。

3 结论与讨论

近年来植物 *CesA* 基因的研究获得了可喜的成绩,但各 *CesA* 基因的精细功能和作用机制都还不清楚。通过生物信息学的比较分析可以预测 *CesA* 基因的作用原理,推测基因家族进化的快慢以及机制,这样可以减少试验的工作量并减少试验的盲目性。

CesA 基因是在转录水平上对纤维素的合成进行调控,在植物细胞壁合成过程中 *CesA* 通过参与纤维素的合成从而发挥重要作用。比较 *CesA* 基因的基因组序列和编码序列,21 个 *CesA* 基因的长度在 3.5~6.5 kb,内含子数量为 7~13 个(*OsCesA10* 除外),*CesA* 基因最大的差异在于某些区域中含子的有无,内含子的位置是确定纤维素合成酶功能的关键因素之一^[26]。

在植株幼根、幼叶和愈伤组织进行组织特异性表达分析发现,水稻和拟南芥中 *CesA* 基因的表达情况大不相同,拟南芥中大部分 *CesA* 基因在幼根、幼叶和愈伤组织中均表达,而水稻大部分 *CesA* 基因则在幼根或幼叶中表达,在愈伤组织中表达的很少。同一个基因在不同组织中表达模式不同,表达量也差别很大,不同物种之间的差别就更大。这些家族成员如何合作来调控代谢途径以及参与植物细胞壁的合成过程还需要进一步研究。

参考文献:

[1] Pauly M, Keegstra K. Cell-wall carbohydrates and their modification as a resource for biofuels[J]. The Plant Journal,2008,54(4):559-568.

[2] Delmer D P. Cellulose biosynthesis: Exciting times for a difficult field of study[J]. Annu Rev Plant Physiol Plant Mol Biol,1999,50:245-276.

[3] Arooli T, peng L, Betzner A S, et al. Molecular analysis of cellulose biosynthesis in *Arabidopsis* [J]. Science, 1998, 279:717-720.

[4] PearJ R, Kawagoe Y, Schreckengost W E, et al. Higher plants contain homologs of the bacterial *CesA* genes encoding the catalytic subunit of cellulose synthase[J]. Proc Natl Acad Sci USA,1996,93:12637-12642.

[5] Doblin M S, Kurek I, Jacob W D, et al. Cellulose biosynthesis in plants: From genes to rosettes [J]. Plant Cell Physiol,2002,43(12):1407-1420.

[6] Holland N, Holland D, Helentjaris T, et al. A comparative analysis of the plant cellulose synthase(*CesA*) gene family [J]. Plant Physiol,2000,123(4):1313-1323.

[7] Appenzeller L, Doblin M, Barreiro R, et al. Cellulose synthesis in maize: Isolation and expression analysis of the cellulose synthase (*CesA*) gene family [J]. Cellulose, 2004,11:287-299.

[8] Burton R A, Shirley N J, King B J, et al. The *CesA* gene family of barley (*Hordeum vulgare*): Quantitative analysis of transcripts reveals two groups of co-expressed genes [J]. Plant Physiol,2004,134:224-236.

[9] Joshi C P, Bhandari S, Ranjan P, et al. Genomics of cellu-

- lose biosynthesis in poplars[J]. *New Phytol*,2004,164:53-61.
- [10] Djerbi S,Lindskog M,Arvestad L,*et al.* The genome sequence of black cottonwood (*Populus trichocarpa*) reveals 18 conserved cellulose synthase(*CesA*) genes[J]. *Planta*,2005,221:739-746.
- [11] Nairn C J,Haselkorn T. Three loblolly pine *CesA* genes expressed in developing xylem are orthologous to secondary cell wall *CesA* genes of angiosperms[J]. *New Phytol*,2005,166:907-915.
- [12] Wang L Q,Guo K,Li Y,*et al.* Expression profiling and integrative analysis of the *CesA/CSL* superfamily in rice [J]. *BMC Plant Biol*,2010,10:282-297.
- [13] Yin Y B,Huang J L,Xu Y. The cellulose synthase superfamily in fully sequenced plants and algae [J]. *BMC Plant Biol*,2009,9:99-113.
- [14] Sundari B K,Dasgupta M G. Isolation of developing secondary xylem specific cellulose synthase genes and their expression profiles during hormone signalling in *Eucalyptus tereticornis*[J]. *J Genet*,2014,93(2):403-414.
- [15] Guerriero G,Legay S,Hausman J F. Alfalfa cellulose synthase gene expression under abiotic stress: A hitchhiker's guide to RT-qPCR normalization[J]. *PLoS One*,2014,9(8):e103808.
- [16] 郭安源,朱其慧,陈新,等. GSDS: 基因结构显示系统 [J]. *遗传*,2007,29(8):1023-1026.
- [17] Sonnhammer E L, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences [J]. *Proc Int Conf Intell Syst Mol Biol*,1998,6:175-182.
- [18] Krogh A, Larsson B, von H G, *et al.* Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes[J]. *J Mol Biol*,2001,305:567-580.
- [19] Horton P, Park K J, Obayashi T, *et al.* WoLF PSORT: Protein localization predictor [J]. *Nucleic Acids Res*,2007,35:585-587.
- [20] Bach I. The LIM domain: Regulation by association [J]. *Mech Dev*,2000,91:5-17.
- [21] Roberts A W, Bushoven J T. The cellulose synthase (*CesA*) gene superfamily of the moss *Physcomitrella patens* [J]. *Plant Mol Biol*,2007,63(2):207-219.
- [22] Hamann T, Osborne E, Youngs H L, *et al.* Global expression analysis of *CesA* and *CSL* gene in *Arabidopsis* [J]. *Cellulose*,2004,11(3):279-286.
- [23] Roberts A W, Roberts E. Cellulose synthase (*CesA*) genes in algae and seedless plant [J]. *Cellulose*,2004,11(3):419-435.
- [24] MacKinnon I M, Sturcova A, Sugimoto S K, *et al.* Cell-wall structure and anisotropy in procuste, a cellulose synthase mutant of *Arabidopsis thaliana* [J]. *Planta*,2006,224(2):438-448.
- [25] Haigler C H, Datcheva I, Hogan P S, *et al.* Carbon partitioning to cellulose synthesis [J]. *Plant Mol Biol*,2001,47(1):29-51.
- [26] Richmond T. Higher plant cellulose synthases [J]. *Genome Biol*,2000,1(4):3001-3006.