

利用全测序基因组数据对被子植物 纤维素合成酶基因的进化分析

王颖, 杨鹏, 于娅, 赵冰, 郭仰东
(中国农业大学农学与生物技术学院, 北京 100193)

摘要:利用已经分离的植物纤维素合成酶基因的 Cellulose_synt 结构域为检索序列,从 NCBI 和其他数据库中调取已完成测序的物种的纤维素合成酶的氨基酸序列,共涉及 10 个物种的 171 个基因,基于以上氨基酸序列,应用 MEGA4.0 生成系统进化树。结果表明: *CesA* 基因和 *Csl* 基因直向的相似度远大于平行的相似度,且它们的分化可能在单子叶和真双子叶植物分化之前,单子叶和真双子叶植物的最近共同祖先中至少存在 7 个 *CesA* 基因,综合已知的模式植物 *CesA* 基因的功能(初生壁或次生壁形成特异性),可为推测其他物种中该基因的功能提供帮助。

关键词:纤维素合成酶基因;被子植物;基因组;系统发育分析

Genome-Wide Comparative Analysis of Angiosperm Cellulose Synthase Gene

WANG Ying, YANG Peng, YU Ya, ZHAO Bing, GUO Yang-dong
(College of Agronomy and Biotechnology, China Agricultural University, Beijing 100193)

Abstract: The cellulose synthase genes were searched and analyzed by using the Cellulose_synt domain sequences identified from the *CesA* genes in plants as queries in NCBI and other genome databases. Phylogenetic trees of cellulose synthase gene from 171 sequences of *CesA* in 10 species including eudicot and monocot published on the GenBank were generated by MEGA4.0. The results showed that these cluster in groups wherein orthologs were often more similar than paralogs, indicating that different subclasses evolved prior to the divergence of the monocot and eudicot lineages. These *CesA* genes in eudicot and monocot were originated from seven genes. Based on the information of biosynthesis of primary or secondary wall by *CesA* gene regulation in model plant, the function of some unknown genes could be speculated from the trees.

Key words: Cellulose synthase gene; Angiosperm; Genome; Phylogenetic analysis

纤维素是细胞壁的主要组成成分,其基本结构单位是吡喃式 D-葡萄糖,以 β -1,4 糖苷键相连^[1]。同时纤维素也是自然界中最丰富的天然高分子有机化合物,每年大约有 1800 亿 t 的纤维素产物生成,巨大的经济价值使纤维素合成酶基因的研究成为热点^[2]。

纤维素合成酶基因编码纤维素合成酶的催化亚基(cellulose synthase subunits),催化合成 β -1,4 糖苷键,是一种重要的糖苷转移酶^[3]。目前很多实验证据表明,纤维素的生物合成是由纤维素合成酶亚基(*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。自然界中除了植

收稿日期:2009-02-25

修回日期:2009-08-31

基金项目:国家重点基础研究发展计划(2009CB119000);国家科技支撑计划(2007BAD57B04);教育部新世纪优秀人才支持计划(NCET-05-0133)

作者简介:王颖,在读硕士,主要从事植物分子生物学研究。E-mail: minnie333@163.com

通讯作者:赵冰,副教授,主要从事蔬菜遗传育种研究。E-mail: zhaocau@163.com

郭仰东,教授,博士生导师,主要从事植物遗传育种与生物技术研究。E-mail: yaguo@cau.edu.cn

物外,很多细菌、真菌和少量的低等动物都能合成纤维素。纤维素合成酶基因首先在木醋杆菌中发现^[4-5],1996年 Delmer 小组从棉花中克隆出第一个植物纤维素合成酶基因^[6],证实了植物中也存在编码纤维合成酶的基因。通过同源性比较发现,这些序列都包含 β -糖苷转移酶的保守氨基酸序列,比较典型的如 D、D、D、QxxRW (D:天冬氨酸,Q:谷氨酰胺,R:精氨酸,W:色氨酸)^[7-8]。

目前对于纤维素合成酶基因的研究多集中在模式植物和经济作物上。有研究表明,拟南芥中除了10个被验证为纤维素合成酶的编码基因外,还有很多基因编码产物也与 AtCesA 具结构上的相似性。由于它们的功能目前还不确定,因而命名为纤维素合成酶相似蛋白 (cellulose-synthase-like protein, Csl 蛋白)。CesA 蛋白和 Csl 蛋白都是糖基转移酶 (GT) 家族 2 中的主要成员,其中纤维素合成酶被列为单独的一族,命名为 CesA。而其余的 Csl 蛋白依据序列结构特征的不同,被分为 6 个亚族,分别命名为 CslA、CslB、CslC、CslD、CslE 与 CslG,它们共同构成纤维素合成酶超基因家族^[9]。

由于现在纤维素合成酶基因的确切功能和不同的纤维素合成酶基因的系统发生关系尚未研究清楚,尤其在国内外针对该基因的研究还是局限于拟南芥和经济作物,很少有关于纤维素合成酶基因进化历史的全面报道。在植物中对该基因的进化研究方面,Noble 等^[10]利用 X 射线、纤维素生物水解标签等方法探讨了蓝藻与植物的内共生可能成为维管植物纤维素合成酶的起源,李益等^[11]选取了 NCBI 中已命名的 15 个物种的 89 个 CesA 基因序列重建了系统发生关系树,通过聚类分析预测了被子植物、裸子植物、苔藓和蕨类植物的 CesA 基因在初生壁和次生壁合成上的特异性,并综合已知的模式植物 CesA 基因的功能,推测某些未知功能基因的可能功能,还发现 CesA 一级结构可变区中半胱氨酸、丙氨酸、甘氨酸和丝氨酸等的含量变化较大,在同一位点半胱氨酸多是与丙氨酸、丝氨酸和缬氨酸发生相互替换,但并未就植物中纤维素合成酶基因的进化历程做出分析。

鉴于先前该基因在被子植物中的进化历史没有被很好地阐明主要是由于数据的缺乏,本研究中我们基于一些已测序的植物的全基因组信息,外加一些具有重要农业经济价值的被子植物的纤维素合成酶基因的氨基酸序列进行更为系统的进化分析,目前在国内未见相关报道,通过构建系统

进化树分析该基因在被子植物内的进化历史,为今后进一步研究该家族的功能和调控机制提供一定的理论依据。

1 材料与方法

1.1 数据的获取

从 Pfam^[12] 网站上下载 Cellulose_synt 结构域 (PF03552) 的 ls 模型,通过 HMMER2.3.2 的 hmmsearch 功能(默认参数)搜索下载到本地的拟南芥 (*Arabidopsis thaliana*; <http://www.arabidopsis.org/>)、水稻 (*Oryza sativa*; <http://rice.plantbiology.msu.edu/>)、毛果杨 (*Populus trichocarpa*; http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html)、葡萄 (*Vitis vinifera*; <http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/>) 蛋白质数据库中含有该结构域的蛋白序列,凡是阈值 $E \leq 0.001$ 的序列都认为是该超基因家族的成员。

在 NCBI 网站和开花植物 EST 库 (The Floral Genome Project, <http://fgp.bio.psu.edu/>) 使用拟南芥的 *AtCesA1*、*AtCesA3*、*AtCesA6*、*AtCesA8* 基因和水稻的 *OsCesA7*、*OsCesA9*、*OsCesA3* 基因的蛋白序列通过 PSI-BLAST 搜索大白菜、油菜、棉花、玉米、小麦、马铃薯 CesA 基因的蛋白序列,凡是阈值 $E \leq 1e^{-3}$ 的序列都认为是 CesA 蛋白序列。

1.2 序列比对与分析

在得到的 10 个物种的 Cellulose_synt 超基因家族的氨基酸序列后,首先利用 Clustal X1.83^[13] 软件对该蛋白家族的氨基酸序列进行多序列比对,参数为默认参数。然后在此基础上进行手工校对修正结构域两侧的比对结果,根据 Pfam 提供的结构域模型去除结构域两端的非同源序列,保留包含结构域的蛋白序列,最后去除所有 gap 并用 MUSCLE3.6^[14] 软件(默认参数)对其重新进行序列比对。依据比对结果分析 Cellulose_synt 结构域的分布特点与排列方式,统计蛋白序列中的保守位点、变异区等参数。

1.3 构建系统发育树

采用 MEGA4.0 软件^[15] 的邻接法 (neighbor-joining tree) 构建邻接树:模式 Amino p-distance,状态 Pairwise Deletion,Bootstrap 检验使用 1000 次重复。

2 结果与分析

2.1 被子植物 CesA 基因的鉴定

在拟南芥、水稻、毛果杨、葡萄基因组中分别筛选到 26、30、39、50 个包含 Cellulose_synt 结构域的

蛋白序列,其中部分已确定功能并加以命名。来源于 NCBI 数据库的大白菜、油菜、棉花、玉米、小麦、马铃薯中已经分离出的 *CesA* 基因及其编码的蛋白质序列共有 26 个(表 1)。

表 1 纤维素合成酶基因的氨基酸序列

Table 1 Amino acid sequences of cellulose synthase gene

植物 Plant	物种 Organism	序列数目 The number of sequences	基因 Gene	数据库蛋白序列编号 Protein No. in database		
单子叶	水稻	30	<i>OsCesA1 ~ 9</i>	Os10g32980.1, Os07g24190.1		
			<i>OsCslD1 ~ 5</i>	Os07g24190.2, Os06g02180.1		
			<i>OsCslE1</i>	Os07g14850.1, Os03g62090.1		
			<i>OsCslE2</i>	Os03g59340.1, Os07g10770.1		
			<i>OsCslE6</i>	Os05g08370.1, Os09g25490.1		
			<i>OsCslF1</i>	Os01g54620.1, Os10g42750.1		
			<i>OsCslF2</i>	Os12g36890.1, Os06g22980.1		
			<i>OsCslF3</i>	Os08g25710.1, Os07g36690.1		
			<i>OsCslF9</i>	Os07g36700.1, Os07g36610.1		
			<i>OsCslH1</i>	Os10g20260.1, Os07g36750.1		
			<i>OsCslH2</i>	Os10g20090.1, Os09g30120.1, Os09g30130.1, Os04g35020.1, Os02g49332.1, Os09g30130.2, Os10g20090.2, Os10g20090.3, Os09g30120.2, Os02g49332.2		
			真双子叶	玉米	12	<i>ZmCesA1 ~ 12</i>
小麦	2	<i>TaCslF6</i>		CAN84874.1, BAD06322.1		
真双子叶	拟南芥	26	<i>AtCesA1 ~ 10</i>	AT5G44030.1, AT5G05170.1		
			<i>AtCslB1 ~ 6</i>	AT4G32410.1, AT1G02730.1		
			<i>AtCslD1 ~ 6</i>	AT4G39350.1, AT3G03050.1		
			<i>AtCslG1 ~ 3</i>	AT5G17420.1, AT4G18780.1		
			<i>AtCslE1</i>	AT2G33100.1, AT5G16910.1, AT4G38190.1, AT5G09870.1, AT2G21770.1, AT5G64740.1, AT2G25540.1, AT1G32180.1, AT1G55850.1, AT2G32620.1, AT2G32530.1, AT4G24010.1, AT2G32540.1, AT4G23990.1, AT4G15290.1, AT2G32610.1, AT4G24000.1, AT4G15320.1		
			毛果杨	39		Pt_176923, Pt_207792, Pt_235238, Pt_245675, Pt_246659, Pt_262611, Pt_286087, Pt_413993, Pt_417275, Pt_551308, Pt_552489, Pt_553321, Pt_554513, Pt_555650, Pt_560094, Pt_560520, Pt_573858, Pt_576348, Pt_595034, Pt_706420, Pt_717644, Pt_750839, Pt_757216, Pt_757958 Pt_763479, Pt_766958, Pt_769203, Pt_784751, Pt_798164, Pt_806784, Pt_813340, Pt_816796, Pt_818594, Pt_819877, Pt_821409, Pt_821920, Pt_827510, Pt_829022, Pt_835809

续表

植物 Plant	物种 Organism	序列数目 The number of sequences	基因 Gene	数据库蛋白序列编号 Protein No. in database
	葡萄	50		Vv_00002977001, Vv_00006699001, Vv_00006701001, Vv_00006702001, Vv_00006704001, Vv_00006705001, Vv_00006707001, Vv_00006708001, Vv_00006709001, Vv_00006710001, Vv_00006712001, Vv_00006715001, Vv_00006717001, Vv_00006719001, Vv_00006720001, Vv_00010208001, Vv_00010942001, Vv_00010944001, Vv_00010945001, Vv_00010946001, Vv_00010947001, Vv_00010948001, Vv_00010950001, Vv_00014029001, Vv_00015671001, Vv_00017182001, Vv_00019341001, Vv_00020164001, Vv_00020168001, Vv_00020169001, Vv_00025577001, Vv_00025901001, Vv_00025903001, Vv_00025904001, Vv_00025905001, Vv_00025906001, Vv_00025909001, Vv_00025911001, Vv_00025913001, Vv_00028062001, Vv_00028523001, Vv_00028800001, Vv_00031177001, Vv_00031746001, Vv_00032621001, Vv_00033394001, Vv_00033397001, Vv_00035840001, Vv_00036303001, Vv_00036318001
	棉花	5	<i>GhCelA1</i> <i>GhCesA1</i> <i>GhCesA3</i> <i>GhCesA4</i> <i>GhCelA2</i>	AAB37766.1 ABC06122.1 AAD39534.2 AAL37718.1 AAB37767.1
	马铃薯	4	<i>StCesA1</i> <i>StCesA3</i> <i>StCesA4</i> <i>StCesA2</i>	AAP97494.1 AAP97495.1 AAP97496.1 AAP97497.1
	大白菜	2		Q4ABM9 Q4ABU5
	油菜	1	<i>BnCesA</i>	ABI34115.1

2.2 被子植物 *CesA* 基因蛋白结构的比较

将已完成测序的模式植物拟南芥、水稻中的 *CesA* 基因的氨基酸序列进行同源性比较,结果发现它们都具有 Cellulose_synt 结构域,特定的锌指结构域和 2 个植物特有的高变区。Cellulose_synt 结构域包含了 3 个 Asp 残基, QXXRW 的保守序列和 8 个跨膜结构域。经蛋白拓扑结构分析可知, 3 个 Asp 残基紧密相连, 与 QXXRW 的保守序列一起在细胞质膜形成底物结合的活性位点(图 1)。另外,植物的 *CesA* 蛋白的 N 端含有两个保守的锌指结构域,具有 4 个保守的串联的半胱氨酸残基 CxxC(图 2), 序列特征为: Cx₂Cx₁₂FxACx₂Cx₂PxCx₂CxEx₃Gx₃Cx₂C, 目前对这些结构域的功能尚不清楚,但推测此序列可与

DNA 结合,可能与 *CesA* 蛋白各亚基之间的相互作用有关^[3]。

2.3 被子植物代表性物种纤维素合成酶基因的系统进化树分析

为了估计被子植物中 *CesA* 蛋白的进化关系,基于 Cellulose_synt 结构域序列,利用 MUSCLE 软件对纤维素合成酶基因编码的氨基酸序列进行同源性比对,再根据上述 10 个物种的 171 条序列利用 MEGA4.0 构建的 NJ 树(图 3)发现,葡萄在与其他物种分开以后独自重复产生了很多拷贝,推测可能功能冗余,也有可能葡萄的纤维素合成酶基因功能相对特异性更高。同时发现,排除葡萄纤维素合成酶基因后构建的 NJ 树,在树形分布及支持率上并没有很大的

```

*****:*** ** * ** * *****:*****:*****:*****:*** *
AtCesA4 VTE D I L T G F R M H C R G W K S V Y C M P K R P A F K G S A P I N L S D R L H Q V L R W A L G S V E I F F S R H C
OsCESA7 VTE D I L T G F K M H C R G W K S V Y C T P A R A A F K G S A P I N L S D R L H Q V L R W A L G S V E I F M S R H C
AtCesA8 I T E D I L T G F K M H C R G W R S I Y C M P L R P A F K G S A P I N L S D R L H Q V L R W A L G S V E I F L S R H C
OsCESA4 VTE D I L T G F K M H C R G W R S I Y C M P I R P A F K G S A P I N L S D R L H Q V L R W A L G S V E I F L S R H C
AtCesA1 VTE D I L T G F K M H A R G W I S I Y C N P P R P A F K G S A P I N L S D R L N Q V L R W A L G S I E I L L S R H C
AtCesA10 VTE D I L T G F K M H A R G W I S I Y C V P S R P A F K G S A P I N L S D R L N Q V L R W A L G S I E I L L S R H C
OsCESA1 VTE D I L T G F K M H A R G W I S I Y C M P P R P C F K G S A P I N L S D R L N Q V L R W A L G S V E I L L S R H C
OsCESA2 VTE D I L T G F K M H A R G W R S I Y C M P K R P A F K G S A P I N L S D R L N Q V L R W A L G S V E I L F S R H C
OsCESA8 VTE D I L T G F K M H A R G W R S I Y C M P K R P A F K G S A P I N L S D R L N Q V L R W A L G S V E I L F S R H C
AtCesA3 VTE D I L T G F K M H A R G W R S I Y C M P K L P A F K G S A P I N L S D R L N Q V L R W A L G S V E I L F S R H C
AtCesA2 VTE D I L T G F K M H C H G W R S V Y C M P K R A A F K G S A P I N L S D R L H Q V L R W A L G S V E I F L S R H C
AtCesA9 VTE D I L T G F K M H C H G W R S V Y C M P K R A A F K G S A P I N L S D R L H Q V L R W A L G S V E I F L S R H C
AtCesA5 VTE D I L T G F K M H S H G W R S V Y C T P K I P A F K G S A P I N L S D R L H Q V L R W A L G S V E I F L S R H C
AtCesA6 VTE D I L T G F K M H S H G W R S V Y C T P K L A A F K G S A P I N L S D R L H Q V L R W A L G S V E I F L S R H C
OsCESA3 VTE D I L T G F K M H C H G W R S I Y C I P K L P A F K G S A P I N L S D R L H Q V L R W A L G S V E I F F S R H C
OsCESA5 VTE D I L T G F K M H C H G W R S I Y C I P K R A A F K G S A P I N L S D R L H Q V L R W A L G S I E I F F S N H C
OsCESA6 I T E D I L T G F K M H C H G W R S I Y C I P K R P A F K G S A P I N L S D R L H Q V L R W A L G S V E I F F S K H C
AtCesA7 I T E D I L T G F K M H C R G W R S I Y C M P K R P A F K G S A P I N L S D R L N Q V L R W A L G S V E I F F S R H S
OsCESA9 I T E D I L T G F K M H C R G W R S V Y C M P K R A A F K G S A P I N L S D R L N Q V L R W A L G S V E I F F S R H S

```

图 1 拟南芥、水稻 *CesA* 基因编码氨基酸序列保守区 QVLRW 的多重比对结果

Fig. 1 Multiple alignment result of conserved domain QVLRW in plant *CesA* amino acid sequences

箭头所示为保守区 QVLRW → for conserved domain QVLRW

```

* * : * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *
AtCesA4 SSFSAK I C K V C G D E V K D D D N G Q T F V A C H V C V Y P V C K P C Y E Y E R S N G N K C C P Q N T L Y K R H K G S
OsCESA7 K --- E K T C R V O G E E V A A R E D G K P F V A C A E C G F V C K P C Y E Y E R S E G T Q C C P Q N T R Y K R H K G C
AtCesA8 ---- S P I Q N T O G E E I G V K S N G E F F V A C H E C S F P I C K A C L E Y E F K E G R R I C L R C G N P Y ----
OsCESA4 ---- V P P C A A C G D D A ---- H A A C R A C S Y A L C K A C L D E A D A E G R T T C A R C G G E Y G ----
AtCesA1 K N M N G I Q Q I C G D D V G L A E T G D V F V A C N E C A F P V C R P C Y E Y E R K D G T Q C C P Q C K T R F R R H K G S
AtCesA10 K D L N G Q I C Q I C G D D V G L T K T G N V F V A C N E C G F P L C Q S C Y E Y E R K D G S Q C C P Q C K A R F R R H N G S
OsCESA1 K S V N G Q V C Q I C G D T V G S A T G D V F V A C N E C A F P V C R P C Y E Y E R K E G N Q C C P Q C K T R Y K R H K G S
OsCESA2 K Q -- C H V C Q I C G D G V G T A A D G E L F T A C D V G F P V C R P C Y E Y E R K D G S Q A C P Q C K T R Y K R H K G S
OsCESA8 R H G S G Q A Q I C G D G V G T T A E G D V F A A C D V G F P V C R P C Y E Y E R K D G T Q A C P Q C K T K Y K R H K G S
AtCesA3 K N I V P Q T C Q I C S D N V G K T V D G D R F V A C D I C S F P V C R P C Y E Y E R K D G N Q S C P Q C K T R Y K R L K G S
AtCesA2 Q E L S G Q T C Q I G D E I E L T V S S E L F V A C N E C A F P V C R P C Y E Y E R R E G N Q A C P Q C K T R Y K R H K G S
AtCesA9 E E L S G Q T C K I C R D E I E L T D N G E P F I A C N E C A F P T C R P C Y E Y E R R E G N Q A C P Q C G T R Y K R I K G S
AtCesA5 E E L S G Q T C Q I C G D E I E L S V D G E S F V A C N E C A F P V C R P C Y E Y E R R E G N Q S C P Q C K T R Y K R I K G S
AtCesA6 Q E L S G Q T C Q I C R D E I E L T V D G E P F V A C N E C A F P V C R P C Y E Y E R R E G N Q A C P Q C K T R F R R H K G S
OsCESA3 R Q N G Q V C Q I C G D D V G L N P D G E P F V A C N E C A F P V C R D C Y E Y E R R E G T Q N C P Q C K T R F K R L R G C
OsCESA5 K H T N G Q V C Q I C G D D V G L T P D G E P F V A C N E C A F P V C R D C Y E Y E R R E G T Q N C P Q C K T R F K R L K G C
OsCESA6 A A E A K A A Q I C G D D V G E G P D G E P F V A C N E C A F P V C R N C Y D Y E R R E G S Q A C P Q C K T R F K R L K G C
AtCesA7 K N L D G Q F C E I C G D Q I G L T V E G D L F V A C N E C G F P A C R P C Y E Y E R R E G T Q N C P Q C K T R Y K R L R G S
OsCESA9 R A L S G Q V C E I C G D E V G R T V D G D L F V A C N E C G F P V C R P C Y E Y E R R E G T Q N C P Q C K T R Y K R L K G S

```

图 2 拟南芥、水稻 *CesA* 基因编码氨基酸序列锌指结构域的多重比对结果

Fig. 2 Multiple alignment result of zinc finger domain in plant *CesA* amino acid sequences

箭头所示为锌指结构域 → for zinc finger domain

影响,这也说明葡萄中的基因不能为阐明该基因在被子植物中的进化提供重大信息。

利用 MEGA4.0 获得的无根树可分为三大支(图 3),分别为 *CesA* 家族、*Csl* 家族中的 D/F 和 B/E/G/H 6 个亚族,其中 *Csl* D/F 归为一支,*Csl* B/E/G/H 归为另一支。由于聚类为无根树,从图中并不能说明 *CesA* 家族与 *Csl* 家族是否有相同的起源,但是就单子叶、真双子叶植物中的 *CesA* 基因和 *Csl* 基因各自聚类,直向的相似度(不同物种的纤维素合成酶基因)远大于平行(同一物种中的纤维素合成酶基因家族各成员)的相似度,可见 *CesA* 基因和 *Csl* 基因的分化是在单子叶和真双子叶植物分化之前。在 II 内包含了 *CslD* 和 *CslF* 两个亚族,其中 *CslD* 与单子叶植物的 *CslF* 聚成一支的支持率为 92%,说明

它们亲缘关系很近,可能是由同一个基因进化而来,这也证实了在单子叶植物中新发现的 *CslF* 亚族与 *CslD* 亚族有很高同源性的原因。同时,从进化树分析发现 *CslD* 家族被聚成 4 小支,且每支聚类的单子叶和真双子叶的支持率达到 80% 以上,表明在单、双子叶植物最近共同祖先中存在 4 个拷贝。但是, *AtCslD6* 和 *Pt176923* 蛋白序列没有和任何一支聚类在一起,可能是 *Pt176923* 蛋白序列测序不全, *AtCslD6* 蛋白序列同源性较低导致在进化树中的位置异常。

III 内分为 *CslB/E/G/H* 4 个亚族,其中真双子叶植物的 *CslB* 与 *CslH* 聚成一支的支持率为 100% 说明它们起源于同一基因,而 *CslE* 与真双子叶植物的 *CslG* 家族聚成一支的支持率为 99%,且从分支长度上看在 III 大支内, *CslB/CslH* 一支的分支长度大

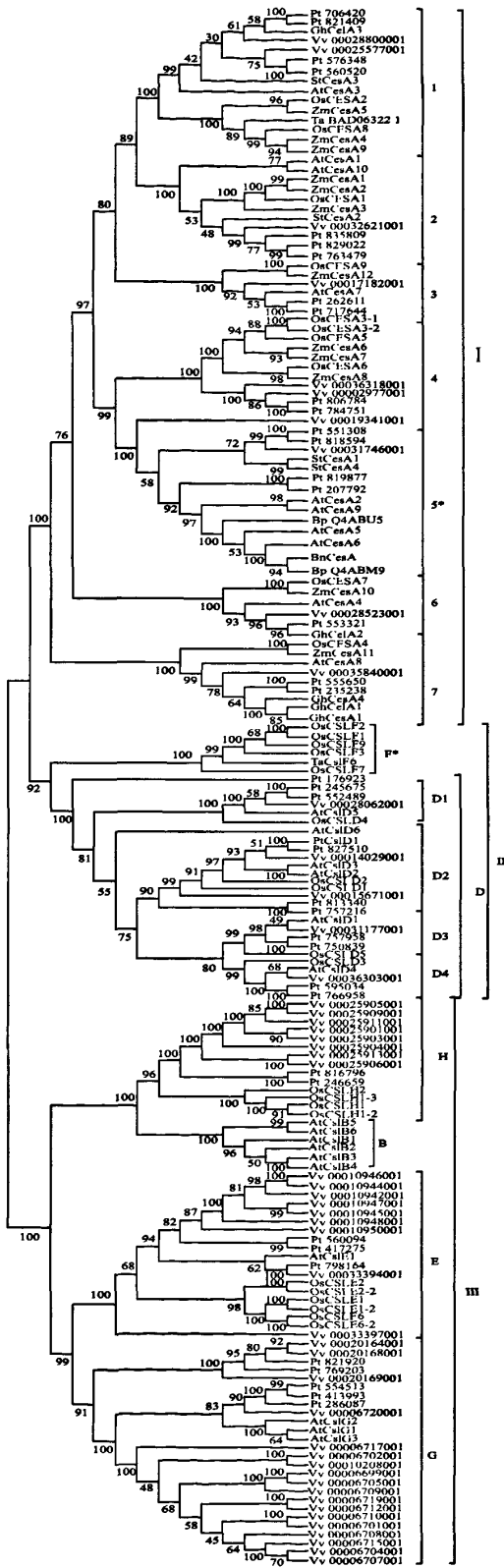


图3 被子植物纤维素合成酶基因蛋白序列的系统发生关系树

Fig.3 Phylogenetic tree of cellulose synthase proteins

于 CslE/CslG 一支的分支长度,提示在 CslB/E/G/H 4 个亚族的进化过程中,各分支上的进化速率并不相同,说明 4 个亚族之间的分化晚于它们的共同祖先和 CslH/CslB 一支的分开。

在 Cesa 家族的一大支 (I) 内分成了 7 个小分支(图 3 的 1~7),据此可推测单、双子叶植物祖先中至少存在 7 个 Cesa 基因的起源,而其中第 5 分支内仅只有真双子叶植物聚类,说明该起源基因在单子叶植物的祖先中丢失了。

以拟南芥、葡萄为例,从单个物种的基因数目来看,目前克隆到的 Csl 家族的序列数多于 Cesa 家族,推测 Csl 家族可能参与细胞壁中其他 β -多聚糖的合成,较 Cesa 基因功能更为复杂,平行变异度更大。而且从已完成基因组测序的物种中发现,每个物种都存在多个的 Cesa 基因和 Csl 基因,可预测在其他物种不仅仅是存在单个的 Cesa 基因和 Csl 基因,也必然是 Cesa 家族和 Csl 家族共同作用植物细胞壁多糖的合成。

2.4 功能预测

根据已知功能的模式植物 Cesa 基因,将图 3 中聚在模式植物基因较密集分支内的未知功能基因进行推测分析。已知模式植物拟南芥 *AtCesa1*、*AtCesa3* 和 *AtCesa6* 在初生细胞壁的纤维素合成中起重要作用,而 *AtCesa4*、*AtCesa7* 和 *AtCesa8* 对于次生细胞壁的纤维素合成是必需的。*OsCesa4*、*OsCesa7* 和 *OsCesa9* 具有细胞次生壁纤维素合成特异性。据此从图 3 中可以看出,单子叶和真双子叶植物的 Cesa 基因大多为初生壁合成特异表达,而且分布比较集中。从而推测具有初生壁特异性的有:*StCesa2*、*StCesa3*、*ZmCesa1*、*ZmCesa2*、*ZmCesa3*、*ZmCesa5*、*ZmCesa4*、*ZmCesa9*、*Ta_BAD06332*、*GhCesa3*、*BnCesa*、*BpQ4ABM9*、*BpABU5*,具有次生壁特异性的有:*ZmCesa12*、*GhCesa1*、*GhCesa2*、*GhCesa4*、*ZmCesa11*、*ZmCesa10*。

3 讨论

目前对纤维素合成酶的分子生物学研究相对比较成熟,从低等的藻类植物、苔藓植物到高等的被子植物都有纤维素合成酶基因鉴定分离出来,用于生化功能的分析;但对于被子植物中纤维素合成酶基因的起源问题尚没有定论。本研究中,我们也对完成基因组测序的藻类—红藻、硅藻、三角褐指藻,利用 Pfam 和 Hmmer 软件进行了 Cellulose_synth 结构域

显示真核生物中的低等生物如红藻、硅藻、三角褐指藻中并不存在任何纤维素合成酶基因。另外,仅在NCBI数据库搜索到掌状冠盖藻中存在2个纤维素合成酶基因(登录号:AF525360,AF525361),同时有研究证实这2个纤维素合成酶基因单独聚类,与其他胚性植物的亲缘关系较远^[16]。苔藓植物中,以获知全基因组序列的小立碗藓为例,将Hmmer2.3.2分析筛选到的16个纤维素合成酶相关基因的蛋白序列构建系统发育树。结果表明小立碗藓仅聚在其中的两小分支内,一支与其他被子植物分布在一起,亲缘关系很近,其余的单独聚在一支,可推测小立碗藓纤维素合成酶基因在进化上可能存在两种起源,一个与维管植物有着共同的祖先,另一个则在维管植物的祖先中丢失了。以掌状冠盖藻、小立碗藓为例与被子植物纤维素合成酶基因的聚类分析结果推测植物界中的藻类、苔藓类与被子植物的纤维素合成酶基因可能不是由相同的基因进化而来,可以说植物中的纤维素合成酶基因的进化历程并没有依循植物由低等到高等的演化规则,这也表明了目前对于维管植物中纤维素合成酶基因的确切起源还有待进一步的研究。

系统发育树的构建是基于多序列比对,依据已知功能的模式植物*CesA*基因,再对聚在同一分支内的未知功能基因进行预测。目前已有研究证明,不同组织或不同细胞壁发育时期,由不同的纤维素合成酶参与合成不同类型的纤维素^[17]。初生、次生细胞壁形成过程中,纤维素的合成涉及不同的纤维素合成酶,这些酶本身的某些特异性序列决定其在初生或次生细胞壁中表达^[18]。本研究中,我们利用Cellulose_synt结构域(PF03552)的1s模型对十字花科中的物种进行了纤维素合成酶家族成员的搜索,结果发现除模式植物拟南芥外,仅在大白菜和油菜中获得3个*CesA*基因的蛋白序列(编号分别为Q4ABM9、Q4ABU5、ABI34115.1),且和*AtCesA6*基因的蛋白序列聚在一起,推测克隆到的基因具有初生壁合成特异性。同时,也可预测到在这两个物种中存在的*CesA*基因多为初生壁特异性,这也为日后十字花科蔬菜中*CesA*基因的功能验证及产品器官纤维含量的调控提供理论依据。但是李益等^[11]证实纤维素合成酶功能的特异性(初生壁与次生壁合成)与该蛋白的一级结构没有直接规律性关联,Kalluri等^[19]从杨树中分离鉴定的木质部中高水平表达基因*PtCesA5*,与拟南芥的*AtCesA3*基因有88%的同

源性,然而功能验证*PtCesA5*基因在木质部发育的次生壁合成期表达,而与其高度同源的*AtCesA3*却在木质部的初生壁合成期表达。因而在分析未知基因的功能时除了以系统进化树信息为理论依据之外,还需要进一步的敲除互补验证。

参考文献

- [1] Arioli T, Peng L, Betzner A S, et al. Molecular analysis of cellulose biosynthesis in *Arabidopsis* [J]. *Science*, 1998, 27: 717-720
- [2] Englehardt J. Sources, industrial derivatives, and commercial applications of cellulose [J]. *Carbohydr Eur*, 1995, 12: 5-14
- [3] Delmer D P. Cellulose biosynthesis: exciting times for a difficult field of study [J]. *Annual Review Plant Physiology and Plant Molecular Biology*, 1999, 50: 245-276
- [4] Saxena I M, Brown R M Jr. Identification of a second cellulose synthase gene (*acsII*) in *Acetobacter xylinum* [J]. *Journal of Bacteriology*, 1995, 177 (18): 5276-5283
- [5] Saxena I M, Lin F C, Brown R M Jr. Identification of a new gene in an operon for cellulose biosynthesis in *Acetobacter xylinum* [J]. *Plant Molecular Biology*, 1991, 16 (6): 947-954
- [6] Pear J, Kawagoe Y, Schrechengost W. Higher plants contain homologs of the bacterial *CelA* genes encoding the catalytic subunit of the cellulose synthase [J]. *PNAS*, 1996, 93: 12637-12642
- [7] Richmond T. Higher plant cellulose synthases [J]. *Genome Biology*, 2000, 1: 1-5
- [8] Brown R M, Saxena I M, Kudlicka K. Cellulose biosynthesis in higher plants [J]. *Trends in Plant Science*, 1996, 5 (1): 149-155
- [9] Somerville C R, Richmond T A. The cellulose synthase superfamily [J]. *Plant Physiol*, 2000, 124: 495-498
- [10] Noble D R, Romanoviza D K, Brown R M. Cellulose in *Cynobacteria* origin of vascular plant cellulose synthase [J]. *Plant Physiol*, 2001, 127: 529-542
- [11] 李益, 胡尚连, 卢学琴, 等. 植物纤维素合成酶基因的进化分析 [J]. *华北农学报*, 2008, 23 (2): 101-105
- [12] Sonnhammer E L, Eoldy S R, Durbin R. Pfam: A comprehensive database of protein domain families based on seed alignments [J]. *Proteins*, 1997, 28 (3): 405-420
- [13] Thompson J D, Gibson T J, Plewniak F, et al. The CLUSTALX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. *Nucleic Acids Research*, 1997, 25 (24): 4876-4882
- [14] Edgar R C. MUSCLE: multiple sequences alignment with high accuracy and high throughput [J]. *Nucleic Acids Research*, 2004, 32 (5): 1792-1797
- [15] Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment [J]. *Brief Bioinformatics*, 2004, 5: 150-163
- [16] Roberts A W, Roberts E M, Delmer D P. Cellulose synthase genes in the green alga *Mesotaenium caldariorum* [J]. *Eukaryotic Cell*, 2002, 1 (6): 847-855
- [17] Haigler C H, Datcheva I, Hogan P S, et al. Carbon partitioning to cellulose synthesis [J]. *Plant Molecular Biology*, 2001, 47: 29-51
- [18] Holland N, Hoiland D, Helenjaris T, et al. A comparative analysis of the plant cellulose synthase gene family [J]. *Plant Physiol*, 2000, 123 (4): 1313-1324
- [19] Kalluri U C, Joshi C P. Isolation and characterization of a new, full-length cellulose synthase cDNA, *PtCesA5* from developing xylem of aspen trees [J]. *J Exp Bot*, 2003, 54: 2187-2188

利用全测序基因组数据对被子植物纤维素合成酶基因的进化分析

作者: [王颖](#), [杨鹏](#), [于娅](#), [赵冰](#), [郭仰东](#), [WANG Ying](#), [YANG Peng](#), [YU Ya](#), [ZHAO Bing](#), [GUO Yang-dong](#)
作者单位: [中国农业大学农学与生物技术学院, 北京, 100193](#)
刊名: [植物遗传资源学报](#) ISTIC|PKU
英文刊名: [JOURNAL OF PLANT GENETIC RESOURCES](#)
年, 卷(期): 2010, 11(2)

参考文献(19条)

- [1. Delmer D P Cellulose biosynthesis:exciting times for a difficult field of study 1999](#)
- [2. Englehardt J Sources, industrial derivatives, and commercial applications of cellulose 1995](#)
- [3. Arioli T;Peng L;Betzner A S Molecular analysis of cellulose biosynthesis in Arabidopsis 1998](#)
- [4. Somerville C R;Richmond T A The cellulose synthase superfamily\[外文期刊\] 2000\(2\)](#)
- [5. Brown R M;Saxenu I M;Kudlicka K Cellulose biosynthesis in higher plants 1996\(I\)](#)
- [6. Richmond T High plant cellulose synthases 2000](#)
- [7. Pear J;Kawagoe Y;Schrechengost W Higher plants contain homologs of the bacterial CelA genes encoding the catalytic subunit of the cellulose synthase 1996](#)
- [8. Saxena I M;Lin F C;Brown R M Jr Identification of a new gene in an operon for cellulose biosynthesis in Acetobacter xylinum 1991\(06\)](#)
- [9. Saxena I M;Brown R M Jr Identification of a second cellulose synthase gene\(acsII\)in Acetobacter xylinum\[外文期刊\] 1995\(18\)](#)
- [10. Kalluri U C;Joshi C P Isolation and characterization of a new, full-length cellulose syhnthase cDNA, PtrCesA5 from developing xylem of aspen trees 2003](#)
- [11. Holland N;Hoiland D;Helenljaris T A comparative analysis of the plant cellulose synthase gene family\[外文期刊\] 2000\(04\)](#)
- [12. Haigler C H;Datcheva I;Hogan P S Carbon partitioning to cellulose synthesis\[外文期刊\] 2001](#)
- [13. Roberts A W;Roberts E M;Delmer D P Cellulose synthase genes in the green alga Mesotaenium culdariorum 2002\(06\)](#)
- [14. Kumar S;Tamura K;Nei M MEGA3:integrated software for molecular evolutionary genetics analysis and sequence alignment 2004](#)
- [15. Edgar R C MUSCLE:multiple sequences alignment with high accuracy and high throushput\[外文期刊\] 2004\(05\)](#)
- [16. Thompson J D;Gibson T J;Plewniak F The CLUSTALX windows interface:Flexible strategies for multiple sequence alignment aided by quality analysis tools\[外文期刊\] 1997\(24\)](#)
- [17. Sonnhammer E L;Eoldy S R;Durbin R Pfam:A comprehensive database of protein domain families based on seed alignments\[外文期刊\] 1997\(03\)](#)
- [18. 李益;胡尚连;卢学琴 植物纤维素合成酶基因的进化分析\[期刊论文\]-华北农学报 2008\(02\)](#)
- [19. Noble D R;Romanoviza D K;Brown B M Cellulose in Cynobacteria origin of vascular plant cellulose synthase 2001](#)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_zwyczyxb201002011.aspx