

# 普通烟草 CESA 基因家族成员的鉴定、亚细胞定位及表达分析

徐宗昌<sup>1,2</sup>, 孔英珍<sup>1</sup>

1. 中国农业科学院烟草研究所, 中国农业科学院烟草遗传改良与生物技术重点开放实验室, 青岛 266101;
2. 中国农业科学院研究生院, 北京 100081

**摘要:** 纤维素合成酶蛋白 (cellulose-synthase proteins, CESA) 是一类质膜定位蛋白, 以蛋白复合体的形式存在于质膜上合成纤维素, 在细胞壁建成和植物生长发育过程中起着非常重要的作用。本研究利用 CESA 蛋白保守域序列 PF03552 检索普通烟草 (*Nicotiana tabacum* L.) 蛋白序列, 并通过拟南芥 (*Arabidopsis thaliana*) 10 个 CESA 蛋白序列在普通烟草基因组数据库中利用 TBLASTN 程序进行比对, 共获得 21 条 *NtCESA* 基因候选序列, 对这些序列进行蛋白序列理化性质分析、系统进化树构建、基因结构分析、保守结构域及跨膜区分析和组织表达模式分析, 并对 *NtCESA9* 和 *NtCESA14* 两个蛋白进行了亚细胞定位实验。结果表明: 获得的 21 条 *NtCESA* 蛋白序列的理化性质相似; 系统进化分析将 21 个 *NtCESA* 基因和 10 个 *AtCESA* 基因分成 5 个分支, 每一个分支各成员之间的进化相对保守, 基因结构类似, 不同分支之间的基因结构差异也较小; *NtCESA* 蛋白结构域相对保守, 都含有 CESA 蛋白典型的 N 端锌指结构、C 端跨膜区和 DDD-QXXRW 保守功能域; 组织表达分析结果表明, 大部分 *NtCESA* 基因在幼苗和成熟期烟草的根、叶、胚芽和愈伤组织中都有表达, 同一个分支中的基因表达模式基本一致, 并且 *NtCESA* 基因参与初/次生细胞壁纤维素的合成与该基因编码蛋白的跨膜区数目存在关联, 表明 *NtCESA* 基因家族成员功能上的复杂性; 亚细胞定位结果证实 *NtCESA9* 和 *NtCESA14* 为质膜定位蛋白。本研究为烟草 CESA 基因家族功能的深入研究奠定了基础。

**关键词:** 普通烟草; CESA 基因家族; 基因表达; 亚细胞定位

## Genome-wide identification, subcellular localization and gene expression analysis of the members of CESA gene family in common tobacco (*Nicotiana tabacum* L.)

Zongchang Xu<sup>1,2</sup>, Yingzhen Kong<sup>1</sup>

1. Key Laboratory of Tobacco Genetic Improvement and Biotechnology, Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao 266101, China;

收稿日期: 2017-02-09; 修回日期: 2017-04-14

基金项目: 国家科技支撑计划项目 (编号: 2015BAD15B03-05) 和国家自然科学基金项目 (编号: 31470291, 31670302) 资助 [Supported by the National Key Technology R&D Program (No.2015BAD15B03-05), and the National Natural Science Foundation of China (Nos.31470291, 31670302)]

作者简介: 徐宗昌, 在读博士研究生, 研究方向: 作物遗传育种。E-mail: xuzc1110@163.com

通讯作者: 孔英珍, 研究员, 博士生导师, 研究方向: 细胞壁多糖合成。E-mail: kongyingzhen@caas.cn

DOI: 10.16288/j.ycz.16-428

网络出版时间: 2017/5/3 16:40:05

URI: <http://kns.cnki.net/kcms/detail/11.1913.R.20170503.1640.002.html>

2. Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, China

**Abstract:** Cellulose-synthase proteins (CESAs) are membrane localized proteins and they form protein complexes to produce cellulose in the plasma membrane. CESA proteins play very important roles in cell wall construction during plant growth and development. In this study, a total of 21 *NtCESA* gene sequences were identified by using PF03552 conserved protein sequence and 10 *AtCESA* protein sequences of *Arabidopsis thaliana* to blast against the common tobacco (*Nicotiana tabacum* L.) genome database with TBLASTN protocol. We analyzed the physical and chemical properties of protein sequences based on some software or on-line analysis tools. The results showed that there were no significant variances in terms of the physical and chemical properties of the 21 *NtCESA* proteins. First, phylogenetic tree analysis showed that 21 *NtCESA* genes and 10 *AtCESA* genes were clustered into five groups, and the gene structures were similar among the genes that are clustered into the same group. Second, in all of the 21 *NtCESA* proteins the conserved zinc finger domain was identified in the N-terminus, transmembrane domains were identified in the C-terminus and the DDD-QXXRW conserved domains were also identified. Third, gene expression analysis results indicated that most *NtCESA* genes were expressed in roots and leaves of seedling or mature tissues of tobacco, seeds and callus tissues. The genes that clustered into the same group share similar expression patterns. Importantly, *NtCESA* proteins that are involved in secondary cell wall cellulose synthesis have two extra transmembrane domains compared with that involved in primary cell wall cellulose biosynthesis. In addition, subcellular localization results showed that *NtCESA9* and *NtCESA14* were two plasma membrane anchored proteins. This study will lay a foundation for further functional characterization of these *NtCESA* genes.

**Keywords:** common tobacco; CESA gene family; gene expression; subcellular localization

纤维素合成酶类似基因家族 CELLULOSE SYNTHASE-LIKE(CSL)属于超基因家族,包含 CSLA、CSLB、CSLC、CSLD、CSLE、CSLF 和 CSLG 7 个亚家族<sup>[1]</sup>。植物细胞壁中的纤维素主要是由 CSLA 基因家族成员编码的纤维素合成酶合成。纤维素是构成植物细胞壁的主要组成成分之一。在植物细胞壁中,纤维素占细胞壁干重的 30%~40% 左右<sup>[2]</sup>,在细胞的生长分化等生理过程中起着重要的作用<sup>[3]</sup>。

在拟南芥(*Arabidopsis thaliana*)中,纤维素合成酶蛋白(cellulose-synthase proteins, CESA)在质膜上合成纤维素<sup>[4]</sup>。每一个 CESA 蛋白都能利用  $\beta$ -D-葡萄糖合成一条没有分支的(1,4)- $\beta$ -D-葡萄糖链,通常每 6 个 CESA 蛋白聚集在一起形成一个复合体,每 6 个复合体聚集到一起形成一个合成纤维素的基本单元,合成一条含有 36 条  $\beta$ -D-葡萄糖链的纤维素微纤丝<sup>[5,6]</sup>,众多的微纤丝构成含量丰富的纤维素。随着研究的深入,越来越多的实验证据表明多个 CESA 蛋白形成的复合体对纤维素的合成是不可或缺的<sup>[5,7,8]</sup>。拟南芥 CESA 蛋白在 N 端有一个保守的锌指蛋白结构域,包含“CXXC”氨基酸残基序列,长度在 10~40 个氨基酸残基左右,形成 CX2CX12FXACX2PXC-

X2CXEX5GX3CX2C 保守域。该保守域能在 CESA 蛋白形成蛋白复合体时进行蛋白之间的相互识别<sup>[9]</sup>。此外,拟南芥 CESA 蛋白具有的保守结构还包括蛋白 C 端的 6 个跨膜区以及 3 个保守的 Asp(D)氨基酸残基<sup>[10]</sup>。在 CESA 蛋白的高级结构中,这 6 个跨膜区能够形成一个分泌葡萄糖链的通道,而 3 个 D 残基则相互毗邻且与另外一个保守结构 QXXRW 靠近,形成一个具有催化糖基转移活性的功能域,与 CESA 蛋白跨膜区形成的葡萄糖链分泌通道相邻<sup>[11]</sup>。

Pear 等<sup>[12]</sup>最早在陆地棉(*Gossypium hirsutum* L.)纤维和水稻(*Oryza sativa* L.)中鉴定到了纤维素合成酶基因。随后,在水稻基因组中共鉴定到 11 个 *OsCESA* 基因,这些基因编码的蛋白包含和拟南芥 CESA 蛋白类似的锌指结构<sup>[13]</sup>。酵母双杂实验表明棉花 CESA 蛋白之间的相互聚合也是依赖蛋白 N 端保守的锌指蛋白结构域<sup>[14]</sup>。在玉米(*Zea mays* L.)基因组中也鉴定到了 12 个 *ZmCESA* 基因,表达实验也证实了 *ZmCesA10-12* 这 3 个基因主要在次生细胞壁中表达<sup>[15]</sup>。目前,关于普通烟草(*Nicotiana tabacum* L.)CESA 基因的研究和报道则相对较少。随着基因组学时代的到来,每年都有大量的植物基因组测序

数据释放,根据这些数据运用相应的分析手段可以对这些物种的 *CESA* 基因进行综合的研究<sup>[16]</sup>。

普通烟草作为特种经济作物,在国民经济中具有重要的地位。同时,普通烟草也是进行生物学研究的重要模式作物,烟草基因组测序的完成<sup>[17]</sup>也为人们在整个基因组水平上进行烟草 *CESA* 基因家族的分析提供了可能。本研究利用已经释放的烟草基因组测序数据,通过同源比对挖掘烟草 *CESA* 基因,构建烟草 *CESA* 基因的系统进化树,并对其基因结构、保守结构域等方面进行了分析,通过亚细胞定位和对烟草转录组数据的分析明确了烟草 *CESA* 基因的亚细胞定位以及组织表达模式,为烟草 *CESA* 蛋白的功能研究奠定了基础。

## 1 材料和方法

### 1.1 材料

本实验所用烟草材料为本生烟(*Nicotiana benthamiana*)和普通烟草品种 K326,由中国农业科学院烟草研究所“国家农作物种质资源平台烟草种质资源子平台”提供。拟南芥为 Columbia 生态型,由本实验室保存。将本生烟、K326 和拟南芥的种子播种于装有蛭石的花盆中,待长出 3 片真叶时,将 K326 幼苗移至装有土的花盆中放于温室中培养;将本生烟和拟南芥幼苗移至装有蛭石的苗盘中放于培养间培养,培养条件为 28℃ 光照 16 h/黑暗 8 h。

收集移栽后 4 周的拟南芥莲座叶、K326 幼苗(转至温室生长约 1~2 周后)以及成株期(开花后植株)植株的根、茎、叶和叶脉等组织(每一个烟草组织样品分别取自长势相似的 3 株烟草植株做为 3 个生物学重复,3 个重复样品进行混合)立即在液氮中速冻,放于-80℃冰箱中保存备用。样品收集完成后立即进行总 RNA 提取,并反转录成 cDNA,用于基因克隆以及定量分析。本生烟在培养间生长约 4~5 周后,挑选具有 4~6 片真叶且长势旺盛的植株进行农杆菌感染。

GFP 融合蛋白表达载体 Pcam35tleGFPs2#4-GFP 和 RFP 融合蛋白表达载体 Pcam35tleGFPs2#4-RFP 为本实验室保存。大肠杆菌菌株 Trans1-T1

购自北京全式金生物技术有限公司,农杆菌感受态 GV3101 由本实验室保存。

### 1.2 普通烟草 *CESA* 基因家族成员的鉴定及序列分析

烟草基因组数据库来源于网站(<https://solgenomics.net/>)。从 Pfam(<http://pfam.xfam.org/>)网站下载 *CESA* 蛋白种子序列(PF03552),利用隐马尔科夫模型软件 HMMER 3.0<sup>[18]</sup>在烟草基因组数据库中进行蛋白序列检索,参数 *E* 值设为 0.001,初步获得烟草 *CESA* 基因家族成员蛋白序列。从拟南芥基因组数据库(<http://www.arabidopsis.org/V10.0>)下载拟南芥的 10 个 *CESA* 蛋白序列<sup>[1]</sup>,分别将这 10 个 *CESA* 蛋白序列在烟草基因组数据库中进行 BLAST 比对,*E* 值设为 0.001。将获得的两部分蛋白序列用 MEGA 6.0<sup>[19]</sup>软件中的 Clustal W 进行多序列比对,剔除重复序列。将上一步所得序列利用 SMART 网站(<http://smart.embl-heidelberg.de/>)进行结构预测,剔除不含 *CESA* 蛋白典型结构域的序列,剩余的蛋白序列视为烟草 *CESA* 基因家族成员。检索家族成员的核酸序列,按文本顺序对烟草 *CESA* 基因按 *NtCESA1*、*NtCESA2*、*NtCESA3* 等的顺序依次命名。采用 ExPaSy 在线软件 ProtParam(<http://web.expasy.org/protparam/>)对该基因家族成员的分子量、理论等电点和脂肪族氨基酸指数等理化性质进行分析。

### 1.3 多序列比对及系统进化树的构建

将拟南芥(10 个)和烟草(21 个)共计 31 个 *CESA* 蛋白序列,用 MEGA 6.0 软件进行多序列比对以及系统进化树的构建。比对的结果采用 Neighbor-Joining(NJ)算法中的 Complete deletion 构建系统进化树,Bootstrap 值取 1000,去除 Bootstrap 支持率低于 50%的节点。

### 1.4 基因结构、蛋白保守结构域及跨膜区预测分析

根据烟草 *CESA* 基因序列和 CDS 序列,利用在线分析工具 GSDS2.0(<http://gsds.cbi.pku.edu.cn/>)<sup>[20]</sup>进行基因结构分析。利用在线工具 MEME(<http://meme-suite.org/tools/meme>)对获得的烟草 *CESA* 基因家族成员的蛋白序列进行保守结构域的预测分析, Motif 具体参数设置为: Motif 长度限制在 6~100 个氨基酸残基范围内;识别的 Motif 总数限制为 7 个。利

用在线分析工具 TMHMM(<http://www.cbs.dtu.dk/services/TMHMM/>)对烟草 CESA 蛋白的跨膜结构进行预测。利用在线预测工具 SWISS-MODEL(<https://swissmodel.expasy.org/>)对烟草 CESA 蛋白的三级结构进行预测。

1.5 烟草 CESA 蛋白的亚细胞定位

1.5.1 烟草 CESA 蛋白亚细胞定位预测

使用在线分析工具 ProtComp 9.0(<http://linux1.softberry.com>)对烟草 CESA 蛋白进行亚细胞定位预测分析。

1.5.2 引物设计

根据数据库中检索到的烟草 CESA 同源基因与拟南芥 CESA 基因的生物信息学分析结果,设计两对带有酶切位点的引物 *NtCESA14-Xba* -F/*NtCESA14-Xba* -R 和 *NtCESA9-Xba* -F/*NtCESA9-Xba* -R 扩增 *NtCESA14* 和 *NtCESA9* 基因;设计 1 对引物 *AtCESA1-Xba* -F/*AtCESA1-Xba* -R, 扩增 *AtCESA1* 基因,做为质膜定位的 Marker 基因。引物序列见表 1。

1.5.3 亚细胞定位表达载体的构建

以普通烟草品种 K326 幼叶和拟南芥莲座叶 cDNA 为模板,用引物 *NtCESA14-Xba* -F/*NtCES-*

*A14-Xba* -R、*NtCESA9-Xba* -F/*NtCESA9-Xba* -R 和 *AtCESA1-Xba* -F/*AtCESA1-Xba* -R 分别对 *NtCESA14*、*NtCESA9* 和 *AtCESA1* 基因进行扩增。PCR 产物经 1.2% 琼脂糖凝胶电泳检测,切胶回收后用 *Xba* 进行单酶切,酶切产物再次纯化。GFP 和 RFP 融合蛋白表达载体 Pcam35tle<sup>g</sup>fps2#4-GFP 和 Pcam-35tle<sup>g</sup>fps2#4-RFP 均使用 *Xba* 进行单酶切,充分酶切后纯化回收表达载体。

将酶切纯化后的 *NtCESA14* 和 *NtCESA9* 基因分别与 GFP 融合蛋白表达载体 Pcam35tle<sup>g</sup>fps2#4-GFP 用 T4 连接酶 4 过夜连接,*AtCESA1* 基因与 RFP 融合蛋白表达载体 Pcam35tle<sup>g</sup>fps2#4-RFP 用 T4 连接酶 4 过夜连接。连接产物进行大肠杆菌转化,将菌落 PCR 检测呈阳性的单菌落进行测序(青岛志远生物公司),序列比对正确后(剔除因单酶切进行反连的克隆)提取质粒,完成表达载体的构建。

1.5.4 亚细胞定位观察

将构建好的表达载体质粒采用电转化法转入农杆菌 GV3101,采用农杆菌浸润法<sup>[21]</sup>注射本生烟叶片,在激光共聚焦显微镜下观察烟草 CESA 融合蛋白 *NtCESA9*-GFP、*NtCESA14*-GFP 和 Marker 蛋白 *AtCESA1*-RFP 在细胞内产生荧光的位置,确定烟草 CESA 蛋白的亚细胞定位情况。

表 1 引物序列信息

Table 1 Primers used in this study

引物名称	序列(5'→3')	T <sub>m</sub> 值( )	扩增长度(bp)
<i>NtCESA14-Xba</i> I -F	TGCTCTAGAAATGAATACTGGAGGGAGGCTT	64	3132
<i>NtCESA14-Xba</i> I -R	TGCTCTAGACTAATCGTCACAATCAATCCG	62	
<i>NtCESA9-Xba</i> I -F	TGCTCTAGAAATGGAGTCAGGGGTTCCGATA	65	2973
<i>NtCESA9-Xba</i> I -R	TGCTCTAGATCAACAATCAATGGAAATGCAGT	62	
<i>AtCESA1-Xba</i> I -F	TGCTCTAGAAATGGAGGCCAGTGCCGGCTTG	70	3243
<i>AtCESA1-Xba</i> I -R	TGCTCTAGACTAAAAGACACCTCCTTTGCCAT	65	
<i>NtCESA16</i> -F	CATCCACCAACCACTCTGA	58	100
<i>NtCESA16</i> -R	GAAGAGAGGACCCCATGAAC	56	
<i>NtCESA19</i> -F	GATCTTCAGTTAGCGTATGTACTTC	55	107
<i>NtCESA19</i> -R	TCAACAACAATAAAGACCACAT	54	
<i>NTubc2</i> -F	CTGACATCTCCCGCACTCTTA	57	179
<i>NTubc2</i> -R	CATAGTCCATTTCGTAGTTGAGCA	56	

注:下划线表示酶切位点。

## 1.6 烟草 CESA 基因表达模式分析

### 1.6.1 转录组表达分析

在 NCBI SRA 数据库(<http://www.ncbi.nlm.nih.gov/sra/>)中下载烟草品种 K326 的转录组测序数据 (SRP029184)。利用 TopHat2<sup>[22]</sup>软件在 K326 转录组数据库中将本研究获得的烟草 CESA 基因家族成员依次检索,提取匹配到的烟草 CESA 基因家族成员的基因表达数据,利用 MeV<sup>[23]</sup>软件将各个基因在不同组织中的表达量数据进行可视化。

### 1.6.2 qRT-PCR 分析

根据多序列比对及系统进化树分析结果,选取 mRNA 序列多态性差异较大的 *NtCESA16* 和 *NtCESA19* 基因做为研究对象,采用 qRT-PCR 分析这两个基因在烟草苗期以及成株期根、茎、叶和叶脉 4 个组织中的表达模式。使用 NCBI 在线分析工

具 Primer-Blast(<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>)在靠近基因 3'端设计定量引物 *NtCESA16-F/NtCESA16-R* 和 *NtCESA19-F/NtCESA19-R*(序列见表 1),扩增片段在 100 bp 左右,以烟草基因 *NtTubc2*(Ubiquitin-conjugating enzyme E2, GenBank 登录号: AB026056)为内参,引物序列见表 1。采用  $2^{-\Delta\Delta Ct}$ <sup>[24]</sup>法计算基因的相对表达量。

## 2 结果与分析

### 2.1 烟草 CESA 基因家族的鉴定及蛋白质理化性质分析

在烟草基因组中进行同源比对,去除重复序列后共鉴定到 21 条烟草 CESA 蛋白序列(表 2)。除 *NtCESA16* 具有 1454 个氨基酸残基外,其余 20 条序列的氨基酸残基数目均在 1000 左右,蛋白大小类似。所有 21 条蛋白序列脂肪族氨基酸指数在 79.39~

表 2 烟草 CESA 基因家族成员鉴定及蛋白理化性质分析

Table 2 The identification and protein physico-chemical analysis of the tobacco CESA gene family

基因编号	氨基酸残基数目	分子量(Da)	理论等电点(pI)	脂肪族氨基酸指数(AI)	疏水性指数(GRAVY)	亚细胞定位	分组	外显子数目
<i>NtCESA1</i>	1085	122049.96	6.50	86.16	-0.21	质膜	Group4	14
<i>NtCESA2</i>	1087	121314.06	7.22	85.80	-0.18	质膜	Group3	14
<i>NtCESA3</i>	1061	119211.77	6.22	87.19	-0.18	质膜	Group4	13
<i>NtCESA4</i>	1085	121975.85	6.48	85.62	-0.21	质膜	Group4	14
<i>NtCESA5</i>	1058	118882.35	6.23	87.07	-0.18	质膜	Group4	13
<i>NtCESA6</i>	1084	122204.15	6.92	86.59	-0.18	质膜	Group1	13
<i>NtCESA7</i>	1041	117813.61	6.46	82.70	-0.20	质膜	Group2	12
<i>NtCESA8</i>	1084	122086.06	6.92	86.85	-0.17	质膜	Group1	13
<i>NtCESA9</i>	991	111472.23	6.24	84.96	-0.11	质膜	Group5	13
<i>NtCESA10</i>	1084	122086.06	6.92	86.85	-0.17	质膜	Group1	13
<i>NtCESA11</i>	1041	117760.56	6.46	83.45	-0.19	质膜	Group2	12
<i>NtCESA12</i>	1087	121308.02	7.22	86.25	-0.18	质膜	Group3	14
<i>NtCESA13</i>	1078	121243.26	6.50	85.13	-0.18	质膜	Group1	13
<i>NtCESA14</i>	1044	117840.15	6.41	84.97	-0.18	质膜	Group1	14
<i>NtCESA15</i>	1091	122761.59	7.65	85.85	-0.19	质膜	Group1	13
<i>NtCESA16</i>	1454	163522.25	6.10	79.39	-0.35	质膜	Group5	17
<i>NtCESA17</i>	1087	122335.14	7.64	86.61	-0.18	质膜	Group1	13
<i>NtCESA18</i>	979	110581.32	6.37	85.90	-0.12	质膜	Group5	14
<i>NtCESA19</i>	932	105345.58	5.68	82.93	-0.18	质膜	Group1	15
<i>NtCESA20</i>	1037	116979.22	6.06	83.84	-0.18	质膜	Group1	14
<i>NtCESA21</i>	1037	116979.22	6.06	83.84	-0.18	质膜	Group1	14



87.19 之间,变化不大;疏水性指数也无太大差异,在-0.35~-0.11 之间。等电点介于 5.68~7.65 之间,其中有 18 个蛋白分子等电点 pH 值为弱酸性,3 个为弱碱性,差异较为明显,这说明绝大多数烟草的 CESA 蛋白表面净电荷为负电荷(表 2)。

## 2.2 系统进化树的构建

烟草的 21 个 CESA 蛋白序列和拟南芥的 10 个 CESA 蛋白序列用 MEGA 6.0 进行系统进化树构建。结果表明,烟草和拟南芥共 31 个 CESA 基因家族成员被划分成了 5 个进化分支(Group1~Group5)(图 1)。在 Group1 中, *AtCESA2*、*AtCESA5*、*AtCESA6* 和 *AtCESA9* 与 *NtCESA6*、*NtCESA8*、*NtCESA10*、*NtCESA13*、*NtCESA14*、*NtCESA15*、*NtCESA17*、*NtCESA19*、*NtCESA20* 及 *NtCESA21* 聚类在一个分支(图 1)。在 Group1 分支中,拟南芥 CESA 基因与烟草 CESA 基因的同源关系较远。*AtCESA7* 和 *NtCESA7*、

*NtCESA11* 聚类在 Group2 分支中; *AtCESA3* 和 *NtCESA2*、*NtCESA12* 聚类在 Group3 分支中; *AtCESA1* 和 *AtCESA10* 与 *NtCESA1*、*NtCESA3*、*NtCESA4* 和 *NtCESA5* 聚类在 Group4 分支中; *AtCESA4*、*AtCESA8* 则和 *NtCESA9*、*NtCESA16* 及 *NtCESA18* 聚类在 Group5 分支中。在 Group2~5 这 4 个分支中,烟草 CESA 基因与拟南芥 CESA 基因同源关系较近(图 1)。

## 2.3 烟草 CESA 基因家族基因结构、蛋白保守结构域及跨膜区预测分析

基因结构分析表明,获得的 21 个烟草 CESA 基因中含有 13 个外显子的基因有 9 个,含有 14 个外显子的基因有 8 个;含有 12 个外显子的基因有 2 个(*NtCESA7* 和 *NtCESA11*), 15 和 17 个外显子的基因各 1 个(分别为 *NtCESA19* 和 *NtCESA16*)(图 2, 表 2)。烟草 CESA 基因家族各成员在外显子数目变化上差异不大。

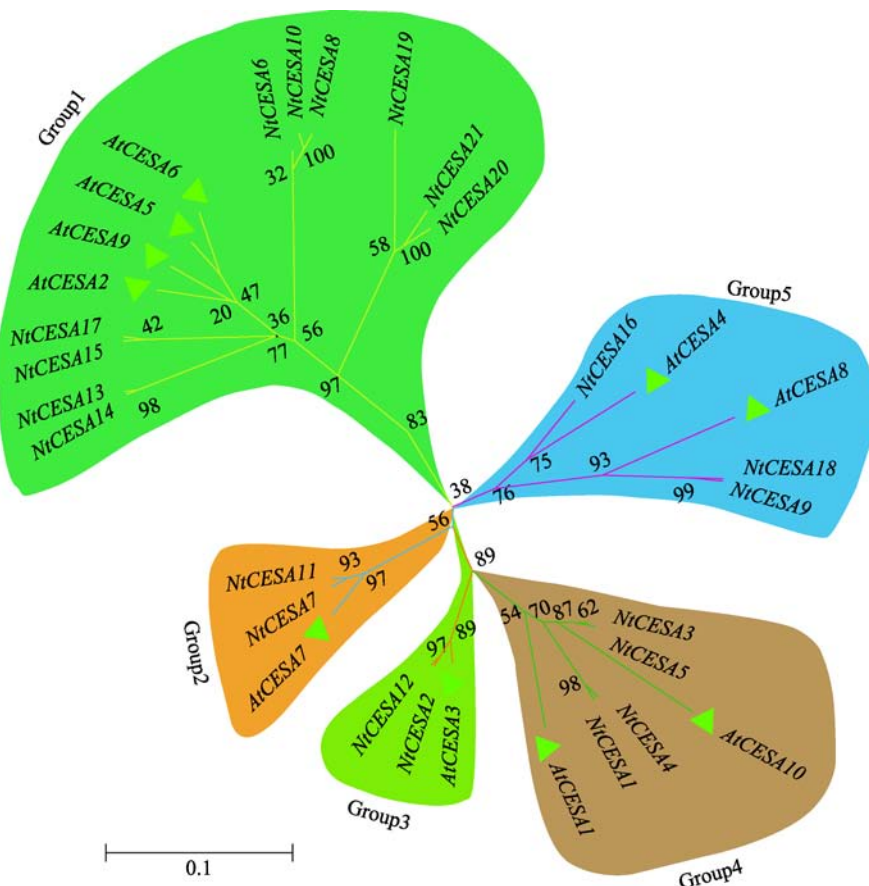


图 1 普通烟草与拟南芥 CESA 基因家族系统进化树分析

Fig. 1 Phylogenetic tree analysis of CESA family genes in *N. tabacum* and *Arabidopsis*

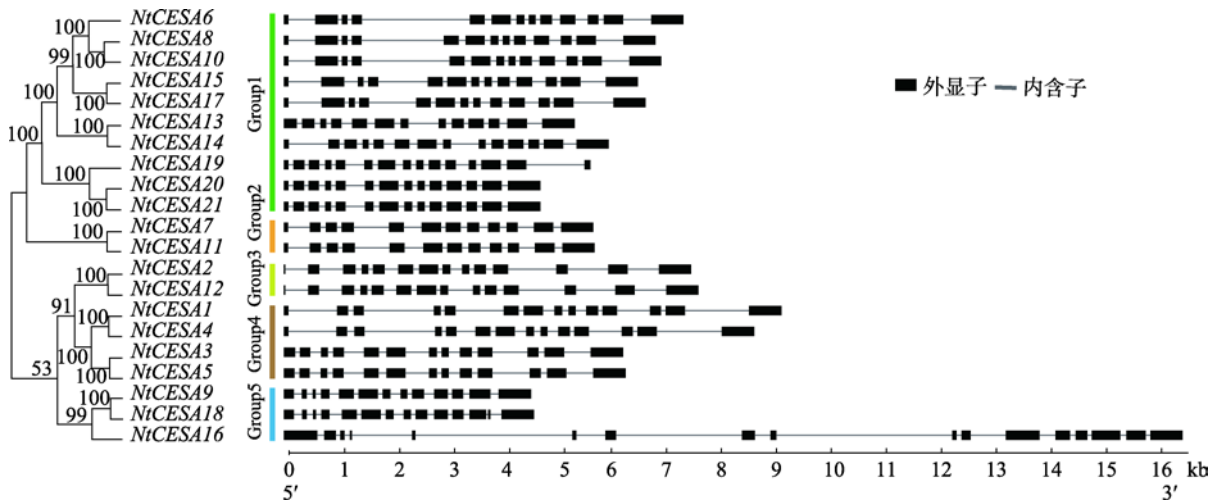


图 2 普通烟草 CESA 基因家族成员基因结构分析

Fig. 2 Gene structure analysis of NtCESA family genes in *N. tabacum*

利用 MEME 在线分析工具对烟草 CESA 基因家族成员进行蛋白序列保守结构域预测分析, 结果表明烟草 CESA 基因家族成员之间蛋白结构高度保守,

烟草 CESA 蛋白序列保守结构域数目和位置基本一致(*NtCESA19* 缺少 Motif2 保守结构域, 只预测到 6 个保守结构域)(图 3), 说明烟草 CESA 蛋白高度保守。

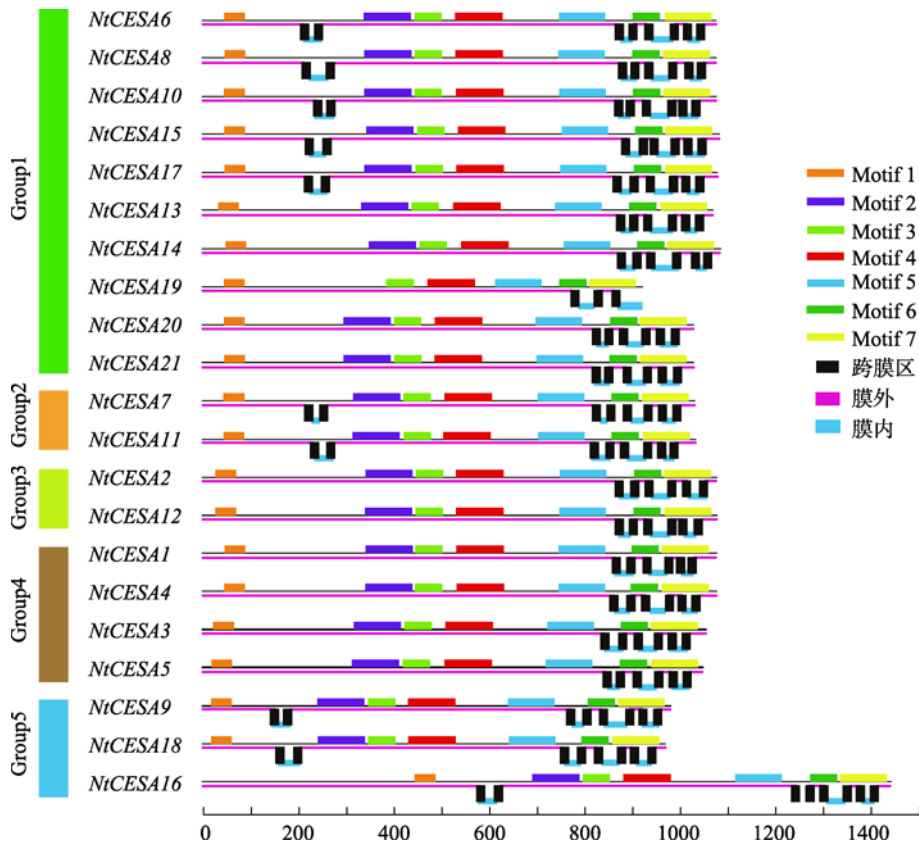


图 3 普通烟草 CESA 基因家族成员蛋白保守结构域及跨膜区预测分析

Fig. 3 Analysis of conserved protein domains and transmembrane domains in members of the NtCESA gene family in *N. tabacum*

所有包含 7 个保守结构域的 20 条蛋白序列中, 在蛋白的 C 端都有 6 个连续的跨膜区, 其中后 5 个跨膜区与 Motif6 和 Motif7 基本重叠(图 3)。大部分烟草 CESA 蛋白都含有这样 6 个跨膜结构, 但是 Group2 和 Group5 中的全部基因及 Group1 中的 5 个基因在靠近 N 端的 Motif1 和 Motif2 保守结构域的中间部分还有两个跨膜结构(图 3)。由于聚类到 Group2 和 Group5 中的烟草 CESA 基因与拟南芥中参与次生细胞壁纤维素合成的基因 *AtCESA4*、*AtCESA7* 和 *AtCESA8*<sup>[25]</sup> 亲缘关系较近, 因此推测这些 *NtCESA* 基因也很可能参与次生细胞壁纤维素的合成。同理, Group3 和 Group4 中的烟草 CESA 基因分别与拟南芥中参与初生细胞壁纤维素合成的基因 *AtCESA3* 和

*AtCESA1*、*AtCESA10*<sup>[26]</sup> 聚类到了同一个分支中, 推测这些 *NtCESAs* 则参与初生细胞壁纤维素的合成。因此, 烟草 CESA 蛋白在 Motif1 和 Motif2 保守结构域处跨膜结构的有无则很可能是参与次/初生细胞壁纤维素合成的关键。

## 2.4 烟草 CESA 蛋白功能结构域分析

对烟草 CESA 蛋白 N 端的保守结构域进行序列比对, 并利用 Weblogo 进行标识, 结果表明烟草 21 条 CESA 蛋白存在一个 “FVACX2CX2PVCX2CX2-YEX3GX3CX2C” 保守功能域(图 4A)。对烟草 CESA 蛋白保守结构域 Motif2~Motif5 进行序列比对, 发现在这几个保守结构域中存在极为保守的功能结构域

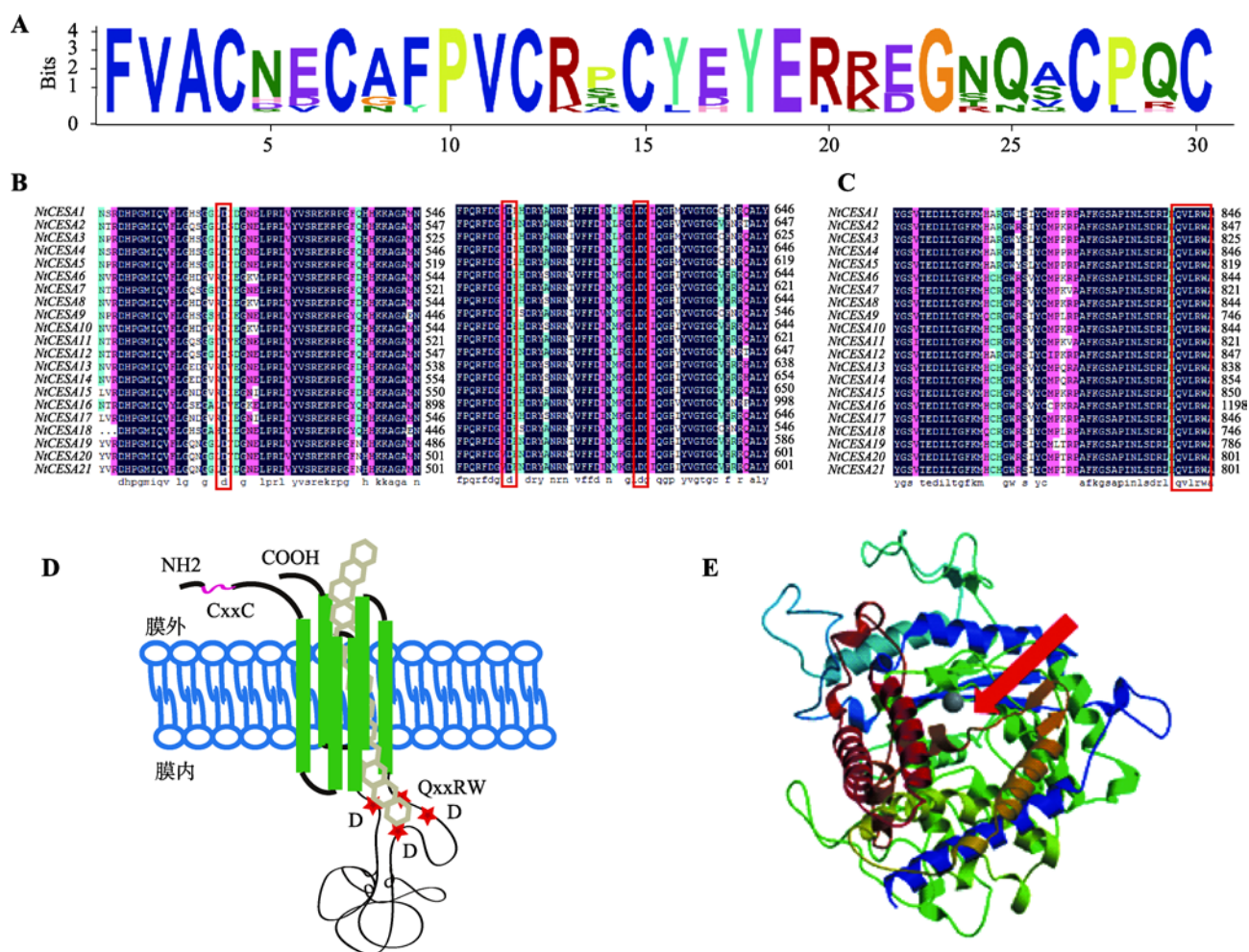


图 4 烟草 CESA 蛋白保守功能域分析及三级结构预测

Fig. 4 The conserved functional domain analysis of *NtCESA* family genes in *N. tabacum*

A: 烟草 *NtCESA* 蛋白 N 端保守锌指蛋白功能域 Weblogo 标识序列; B: 烟草 *NtCESA* 蛋白 D、DXD 保守功能域; C: 烟草 *NtCESA* 蛋白 QXXRW 保守功能域; D: 烟草 *NtCESA* 蛋白功能模型; E: 烟草 CESA 蛋白三级结构预测。



D、DXD(图 4B)和 QXXRW(图 4C), 在 21 条烟草 CESA 蛋白中均存在这些功能域, 这与报道的拟南芥 CESA 蛋白具有相同的功能结构域<sup>[10]</sup>。CESA 蛋白催化模型如图 4D 所示, 跨膜区聚集在膜上形成一个分泌葡萄糖链的通道, 而 3 个 D 残基则相互毗邻且与另外一个保守结构 QXXRW 共同位于 CESA 蛋白靠近分泌通道的活性区域, 催化糖链的合成<sup>[11]</sup>。在烟草 CESA 蛋白的三级结构预测模型中也观察到了这个葡萄糖链通道的存在(图 4E)。烟草 CESA 蛋白的三级结构多以  $\alpha$  螺旋和不规则卷曲形成, 也有少量的  $\beta$  折叠存在。在三级结构中, 跨膜区主要由  $\alpha$  螺旋构成, 多个跨膜区相互勾连交合, 形成一个孔型通道(图 4E 中红色箭头所示)。

## 2.5 烟草 CESA 蛋白亚细胞定位研究

烟草 CESA 蛋白亚细胞定位预测结果表明, 烟草 CESA 蛋白都定位在质膜上(表 2)。通过系统进化树、基因结构分析以及保守功能域发现, 烟草 CESA 基因与拟南芥 CESA 基因有着相似的生物学功能。为了进一步验证预测分析结果, 本文扩增了 *NtCESA9*

和 *NtCESA14* 基因, 并分别将 *NtCESA9* 和 *NtCESA14* 与 GFP 融合蛋白表达载体 Pcam35tlefps2#4-GFP 连接, 构建可瞬时表达 *NtCESA9*-GFP 和 *NtCESA14*-GFP 融合蛋白的载体, 与质膜定位融合蛋白 *AtCESA1*-RFP 进行亚细胞共定位研究, 探索烟草 CESA 蛋白在生物体内的定位情况。

亚细胞定位观察结果表明, 质膜定位的 Marker 融合蛋白 *AtCESA1*-RFP 红色荧光信号主要集中在质膜上, 烟草 CESA 融合蛋白 *NtCESA9*-GFP 和 *NtCESA14*-GFP 的绿色荧光信号也主要集中在质膜上, 并与 *AtCESA1*-RFP 的红色荧光信号基本重合(图 5), 说明 *NtCESA9* 和 *NtCESA14* 两个蛋白定位在质膜上。

## 2.6 烟草 CESA 基因家族表达分析

在普通烟草品种 K326 的转录组数据库中, 通过 TopHat2 软件比对一共检索到了 17 个烟草 CESA 基因家族成员(*NtCESA1*、*NtCESA4*、*NtCESA19* 和 *NtCESA21* 4 个基因除外)的表达量数据。使用 MeV 软件将这些基因在 K326 不同组织的表达水平进行

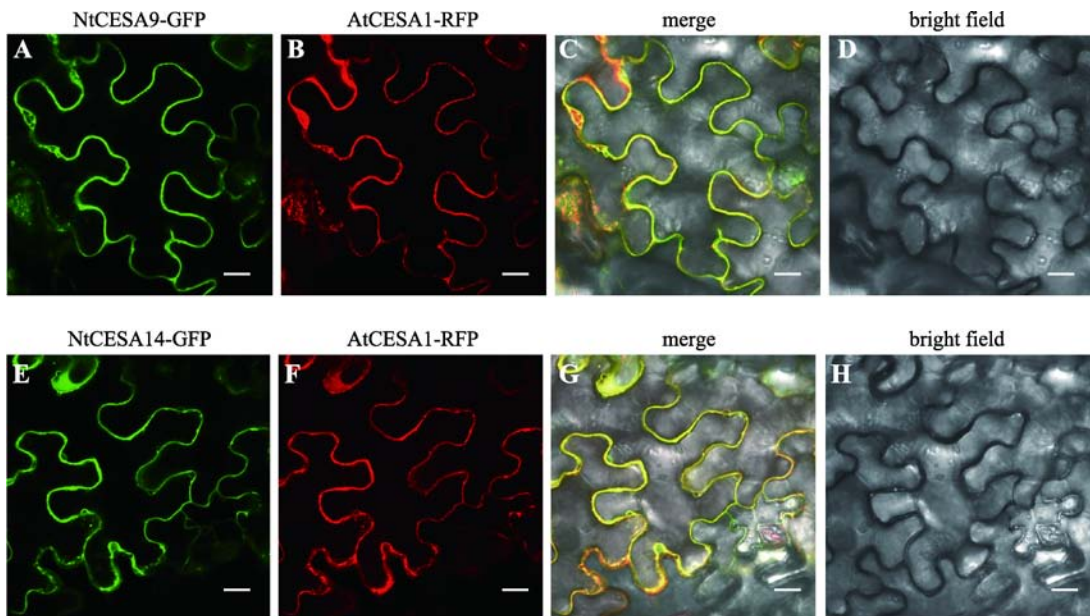


图 5 烟草 *NtCESA9* 和 *NtCESA14* 蛋白亚细胞定位结果

Fig. 5 The subcellular localization results of *NtCESA9* and *NtCESA14* proteins

A: 烟草 *NtCESA9*-GFP 绿色荧光融合蛋白在本生烟叶表皮细胞中的定位; B: 与烟草 *NtCESA9*-GFP 共定位的质膜定位 Marker 蛋白 *AtCESA1*-RFP 红色荧光融合蛋白在本生烟叶表皮细胞中的定位; C: A 与 B 的重叠图像; D: A、B 和 C 视野明场视场; E: 烟草 *NtCESA14*-GFP 绿色荧光融合蛋白在本生烟叶表皮细胞中的定位; F: 与烟草 *NtCESA14*-GFP 共定位的质膜定位 Marker 蛋白 *AtCESA1*-RFP 红色荧光融合蛋白在本生烟叶表皮细胞中的定位; G: E 与 F 的重叠图像; H: E、F 和 G 视野明场视场。绿色荧光激发光为 488 nm, 红色荧光激发光为 590 nm, 标尺=20  $\mu\text{m}$ 。

可视化分析, 结果表明, 处于同一个进化分支的不同基因在多个组织中的表达模式基本一致; 除 *NtCESA20* 在各个组织中几乎没有表达外, 其余 16 个基因在成熟/幼苗期烟草的根、叶、胚芽和愈伤组织中都有表达(图 6)。Group2(*NtCESA7* 和 *NtCESA11*) 和 Group5(*NtCESA9*、*NtCESA16* 和 *NtCESA18*) 中的基因主要集中在幼苗根和成熟期植株的根中表达, 其他组织里的表达量较低, 组织表达的特异性较高。而 Group1(*NtCESA15* 和 *NtCESA17*)、Group3 (*NtCESA2* 和 *NtCESA12*) 和 Group4(*NtCESA3* 和 *NtCESA5*) 中的这些基因在这 7 个组织中都有很高的表达, 没有体现出组织特异性表达的特点(图 6)。

Group2 和 Group5 中烟草 CESA 基因与拟南芥中主要合成次生细胞壁纤维素基因 *AtCESA4*、*AtCESA7* 和 *AtCESA8* 聚类在一个分支中(图 1), 因此本文推测这两个分支中的基因也很有可能参与次生细胞壁纤维素的合成, 并且转录组数据也表明这两个分支中的基因在成熟期植株根中的表达要高于幼苗根中的表达。为了进一步证实推论的准确性, 本文进一步挑选 *NtCESA16* 和 *NtCESA19* 基因, 通过

qTR-PCR 分析这两个基因在烟草苗期和成熟期根、茎、叶和叶脉中的表达情况。结果表明, 相对于成熟期材料, *NtCESA19* 在幼嫩材料的根、叶和叶脉中的表达量显著升高(图 7), 这说明 *NtCESA19* 很可能主要参与初生细胞壁纤维素的合成; 而 *NtCESA16* 在幼苗期材料的茎和叶脉中的表达量要显著低于成熟期材料, 在幼苗期根中的表达略低于成熟期的根, 与转录组数据基本一致, 说明 *NtCESA16* 很可能主要参与次生细胞壁纤维素的合成。

### 3 讨 论

*CESA* 基因编码的纤维素合成酶蛋白能够在质膜上以复合体的形式合成纤维素<sup>[5,7,8]</sup>。纤维素做为植物细胞壁的重要组成部分之一, 不仅参与细胞形态的建成, 控制细胞生长, 而且还能参与到多种生物、非生物胁迫的信号应答调控网络中, 因此在植物的生长发育过程中起着非常重要的作用<sup>[6,27~29]</sup>。

本研究基于已经测序完成的烟草基因组数据<sup>[17]</sup>, 利用生物信息学分析的方法首次从烟草中鉴定到了 21 个烟草 CESA 基因家族成员。该基因家族各成员

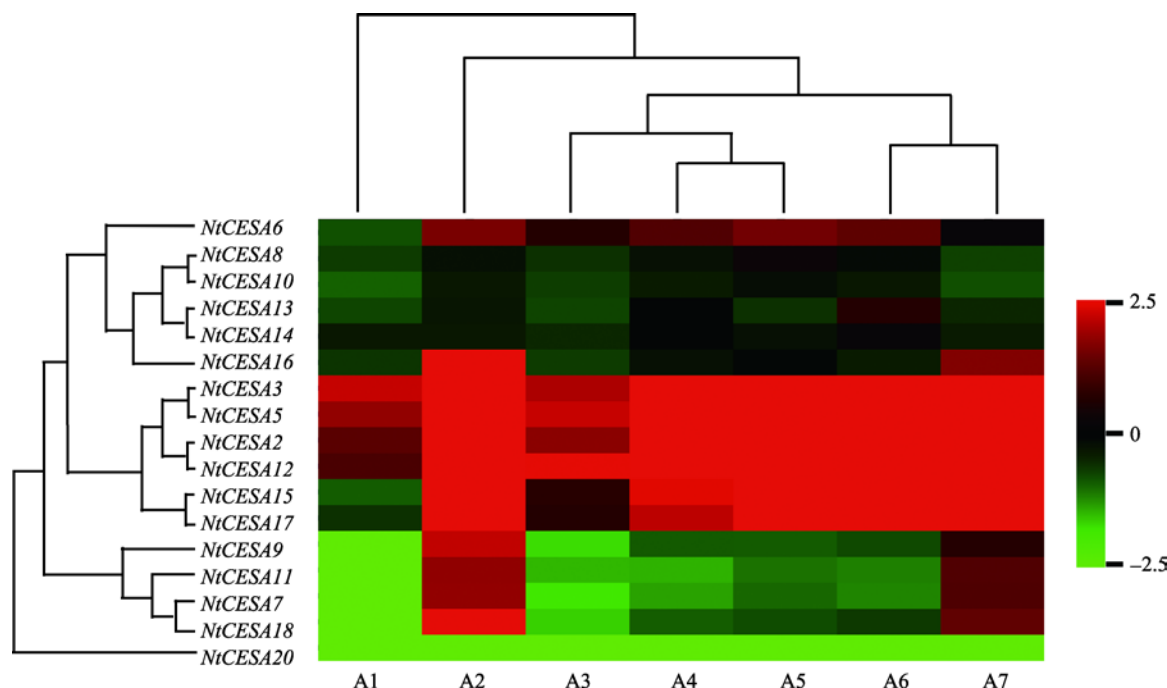


图 6 烟草 CESA 基因家族表达模式分析

Fig. 6 Expression patterns of the members of the tobacco CESA gene family

A1: 成熟期植株叶片; A2: 成熟期植株根; A3: 胚芽; A4: 愈伤组织; A5: 成熟期植株叶脉; A6: 幼苗叶; A7: 幼苗根。红色表示基因表达量高, 绿色表示基因表达量低。

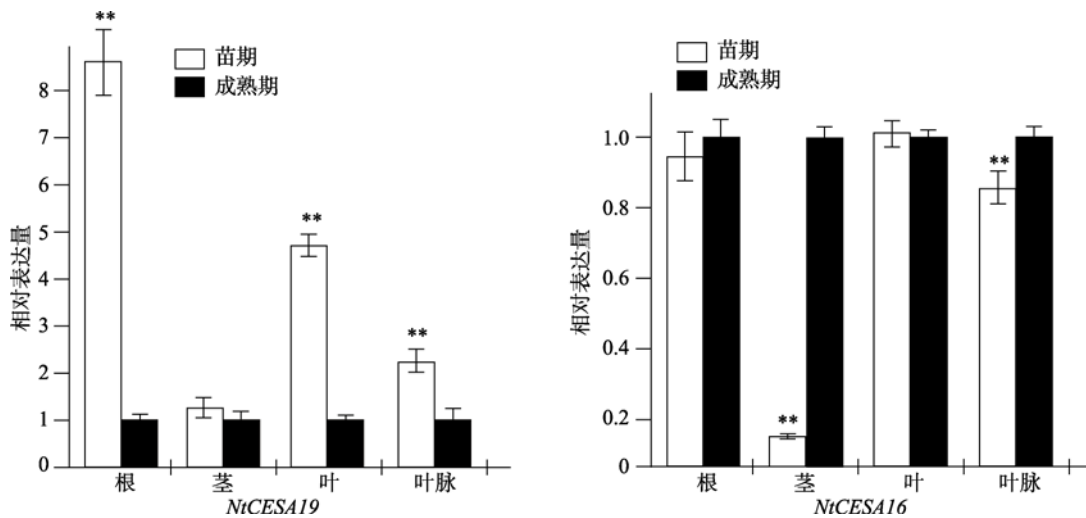


图 7 *NtCESA16* 和 *NtCESA19* 在烟草幼苗及成株根、茎、叶和叶脉中的表达量分析

Fig. 7 qRT-PCR analysis of the expression level of *NtCESA16* and *NtCESA19* in tobacco seedlings and adult plants roots, stems, leaves and leaf veins tissues

误差线表示 3 次独立重复实验平均值的标准差, \*\*表示双尾 *t* 检验条件下, 基因在苗期与成熟期的表达量存在极显著差异,  $P < 0.01$ 。

之间在基因全长、外显子数目、蛋白质长度等方面没有明显差异(图 2, 表 2)。对该基因家族各成员的保守结构域预测结果表明, 每个成员都含有保守的 DDD-QXXRW 氨基酸残基序列(图 4: B, C), 这些保守的氨基酸残基在蛋白质折叠形成三级结构时能够在空间上相互接近, 形成一个葡萄糖链的分泌通道<sup>[11]</sup>(图 4: D, E)。在拟南芥<sup>[10]</sup>和水稻<sup>[13]</sup>的 CESA 蛋白中也发现了相同的保守氨基酸残基, 说明这些保守的氨基酸残基可能是 CESA 蛋白行使正常功能所必须的, 在物种间相对保守。另外, 本文还发现烟草 CESA 蛋白 N 端保守的锌指结构(图 4A)与拟南芥锌指结构<sup>[11]</sup>略有不同, 本文推测这与物种特异性有关, 不同物种 CESA 蛋白形成多聚体时相互识别的序列存在差异。

普通烟草是异源四倍体植株, 基因组中含有 SSTT 4 个染色体基因组( $2n=4x=48$ , SSTT)<sup>[30]</sup>。现在一般认为普通烟草的 S 基因组供体是林烟草(*Nicotiana sylvestris*,  $2n=2x=24$ , SS), T 基因组供体是茸毛状烟草(*Nicotiana tomentosiformis*,  $2n=2x=24$ , SS), 两个祖先种大约在距今 20 万年前经过一次种间杂交产生了普通烟草<sup>[30]</sup>, 因此普通烟草基因组中存在两套部分同源的染色体基因组 SS 和 TT。本文对烟草 CESA 基因家族成员进行系统进化树分析, 发现 Group2 中的 *NtCESA7*、*NtCESA11* 基因和 Group3 中

的 *NtCESA2*、*NtCESA12* 基因; 以及 Group1 中的 *NtCESA6*、*NtCESA8* 和 *NtCESA103* 个基因, *NtCESA15* 和 *NtCESA17* 两个基因, *NtCESA20* 和 *NtCESA21* 两个基因; Group4 中的 *NtCESA1* 和 *NtCESA4* 以及 *NtCESA3* 和 *NtCESA5* 两对基因有着极为一致的基因结构, 外显子的数目以及在基因上的位置基本一致(图 2)。本文推测上述基因结构相似的 *NtCESAs* 很有可能是分别来源于两个不同的祖先种, 他们在长时间的进化过程中没有发生分化或者融合, 但此类基因是否都具有生物学功能仍需要进一步的实验验证。

初生细胞壁是在细胞伸长的过程中合成的, 相对较薄, 容易变化并且含水较高, 植物快速分裂生长的幼嫩组织如幼根、幼叶、幼茎中富含初生细胞壁; 而次生细胞壁则是在细胞伸长停止以后在某一些种类的细胞上沉积而成, 一般而言植物特化和成熟变老的组织如老根、叶脉、成株期植株茎秆的细胞壁富含次生细胞壁<sup>[6]</sup>。有研究报道在拟南芥中, *AtCESA1*、*AtCESA3* 和 *AtCESA6* 主要参与初生细胞壁纤维素的合成<sup>[25]</sup>, 这 3 个基因被聚类到了 Group1、Group3 和 Group4 中(图 1); *AtCESA4*、*AtCESA7* 和 *AtCESA8* 则主要参与次生细胞壁纤维素的合成<sup>[26]</sup>, 它们被聚类到了 Group2 和 Group5 中(图 1)。一般而言, 聚类到同一个进化分支上的基因具有类似的功能, 本文通过在烟草转录组数据库中检索 CESA 基



因家族成员的表达模式发现, Group2 和 Group5 中的烟草 CESA 基因家族成员在成熟期植株根中的表达要比在幼苗根中的表达高(图 6), 说明 *NtCESA7*、*NtCESA9*、*NtCESA11*、*NtCESA16*、*NtCESA18* 等基因主要参与烟草次生细胞壁纤维素的合成。另外, 对烟草 CESA 基因家族成员蛋白跨膜区预测分析发现这 5 个基因编码的 CESA 蛋白具有 8 个跨膜区(图 3), 而与拟南芥中参与初生细胞壁纤维素合成的 *AtCESA1*、*AtCESA3* 和 *AtCESA6* 基因聚类到 Group1、Group3 和 Group4 中的绝大多数烟草 CESA 基因家族成员只有 6 个跨膜区(图 3)。在拟南芥中无论是参与初生细胞壁纤维素合成还是次生细胞壁纤维素合成的 CESA 蛋白都有 8 个跨膜区<sup>[10]</sup>, 因此本文推测烟草中 CESA 蛋白跨膜区数目的差异是该基因参与初生细胞壁或次生细胞壁纤维素合成的一个特征。*NtCESA16* 和 *NtCESA19* 分别具有 8 个和 6 个跨膜区, qRT-PCR 结果证实 *NtCESA16* 主要在成熟期烟草的茎和叶脉中表达, 合成次生细胞壁纤维素; 而 *NtCESA19* 主要在幼苗期烟草的叶和叶脉中表达, 合成初生细胞壁纤维素(图 7), 这与我们的推测一致。本文初步探明了烟草 CESA 蛋白跨膜区数目与参与初生/次生壁合成之间存在关联, 为以后研究烟草 CESA 基因的功能提供了实验基础。

## 参考文献(References):

- [1] Richmond TA, Somerville CR. The cellulose synthase superfamily. *Plant Physiol*, 2000, 124(2): 495–498. [DOI]
- [2] Liepman AH, Wightman R, Geshi N, Turner SR, Scheller HV. Arabidopsis—a powerful model system for plant cell wall research. *Plant J*, 2010, 61(6): 1107–1121. [DOI]
- [3] Cosgrove DJ, Jarvis MC. Comparative structure and biomechanics of plant primary and secondary cell walls. *Front Plant Sci*, 2012, 3: 204. [DOI]
- [4] Kimura S, Laosinchai W, Itoh T, Cui XJ, Linder CR, Brown RM Jr. Immunogold labeling of rosette terminal cellulose-synthesizing complexes in the vascular plant *Vigna angularis*. *Plant Cell*, 1999, 11(11): 2075–2086. [DOI]
- [5] Perrin RM. Cellulose: how many cellulose synthases to make a plant? *Curr Biol*, 2001, 11(6): R213–R216. [DOI]
- [6] Cosgrove DJ. Growth of the plant cell wall. *Nat Rev Mol Cell Biol*, 2005, 6(11): 850–861. [DOI]
- [7] Scheible WR, Eshed R, Richmond T, Delmer D, Somerville C. Modifications of cellulose synthase confer resistance to isoxaben and thiazolidinone herbicides in *Arabidopsis* *Ixr1* mutants. *Proc Natl Acad Sci USA*, 2001, 98(18): 10079–10084. [DOI]
- [8] Taylor NG, Laurie S, Turner SR. Multiple cellulose synthase catalytic subunits are required for cellulose synthesis in Arabidopsis. *Plant Cell*, 2000, 12(12): 2529–2540. [DOI]
- [9] Zhou XF, Wang JY, Wang XZ. Research progress of cellulose synthase genes in higher plant. *Hereditas (Beijing)*, 2002, 24(3): 376–378.  
周晓馥, 王景余, 王兴智. 植物纤维素合成酶基因的研究进展. *遗传*, 2002, 24(3): 376–378. [DOI]
- [10] Delmer DP. Cellulose biosynthesis: exciting times for a difficult field of study. *Annu Rev Plant Physiol Plant Mol Biol*, 1999, 50: 245–276. [DOI]
- [11] Richmond T. Higher plant cellulose synthases. *Genome Biol*, 2000, 1(4): reviews3001.1–3001.6. [DOI]
- [12] Pear JR, Kawagoe Y, Schreckengost WE, Delmer DP, Stalker DM. Higher plants contain homologs of the bacterial *celA* genes encoding the catalytic subunit of cellulose synthase. *Proc Natl Acad Sci USA*, 1996, 93(22): 12637–12642. [DOI]
- [13] Wang LQ, Guo K, Li Y, Tu YY, Hu HZ, Wang BR, Cui XC, Peng LC. Expression profiling and integrative analysis of the *CESA/CSL* superfamily in rice. *BMC Plant Biol*, 2010, 10(1): 282. [DOI]
- [14] Kurek I, Kawagoe Y, Jacob-Wilk D, Doblin M, Delmer D. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proc Natl Acad Sci USA*, 2002, 99(17): 11109–11114. [DOI]
- [15] Appenzeller L, Doblin M, Barreiro R, Wang HY, Niu XM, Kolipara K, Carrigan L, Tomes D, Chapman M, Dhugga KS. Cellulose synthesis in maize: isolation and expression analysis of the cellulose synthase (*CesA*) gene family. *Cellulose*, 2004, 11(3–4): 287–299. [DOI]
- [16] Kumar M, Turner S. Plant cellulose synthesis: CESA proteins crossing kingdoms. *Phytochemistry*, 2015, 112: 91–99. [DOI]
- [17] Sierro N, Battey JND, Ouadi S, Bakaher N, Bovet L, Wilzig A, Goepfert S, Peitsch MC, Ivanov NV. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun*, 2014, 5: 3833. [DOI]
- [18] Finn RD, Clements J, Eddy SR. HMMER web server: in-teractive sequence similarity searching. *Nucleic Acids Res*, 2011, 39(S2): W29–W37. [DOI]
- [19] Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S.



- MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*, 2013, 30(12): 2725–2729. [DOI]
- [20] Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 2015, 31(8): 1296–1297. [DOI]
- [21] Jin YR, Song YF, Bai Y, Zhang L, Dong LH, Liu CK, Feng XG, Hu XM, Wang Q, Liu HB. Molecular cloning and expression analysis of potassium channel gene *NKT6* in *Nicotiana glauca*. *Acta Agronom Sinica*, 2013, 39(9): 1602–1611.  
靳义荣, 宋毓峰, 白岩, 张良, 董连红, 刘朝科, 冯祥国, 胡晓明, 王倩, 刘好宝. 林烟草钾离子通道基因 *NKT6* 的克隆与表达定位分析. 作物学报, 2013, 39(9): 1602–1611. [DOI]
- [22] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, 2013, 14(4): R36. [DOI]
- [23] Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J. TM4: a free, open-source system for microarray data management and analysis. *BioTechniques*, 2003, 34(2): 374–378. [DOI]
- [24] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C_T}$  method. *Methods*, 2001, 25(4): 402–408. [DOI]
- [25] Arioli T, Peng LC, Betzner AS, Burn J, Wittke W, Herth W, Camilleri C, Höfte H, Plazinski J, Birch R, Cork A, Glover J, Redmond J, Williamson RE. Molecular analysis of cellulose biosynthesis in *Arabidopsis*. *Science*, 1998, 279(5351): 717–720. [DOI]
- [26] Taylor NG, Howells RM, Huttly AK, Vickers K, Turner SR. Interactions among three distinct Cesa proteins essential for cellulose synthesis. *Proc Natl Acad Sci USA*, 2003, 100(3): 1450–1455. [DOI]
- [27] Burton RA, Gidley MJ, Fincher GB. Heterogeneity in the chemistry, structure and function of plant cell walls. *Nat Chem Biol*, 2010, 6(10): 724–732. [DOI]
- [28] Caffall KH, Mohnen D. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res*, 2009, 344(14): 1879–1900. [DOI]
- [29] Atmodjo MA, Hao ZY, Mohnen D. Evolving views of pectin biosynthesis. *Annu Rev Plant Biol*, 2013, 64(1): 747–749. [DOI]
- [30] Leitch IJ, Hanson L, Lim KY, Kovarik A, Chase MW, Clarkson JJ, Leitch AR. The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae). *Ann Bot*, 2008, 101(6): 805–814. [DOI]

(责任编辑: 吴为人)