

一个新的棉纤维表达蛋白 cDNA 的克隆、表达及功能初步分析

郭 婕, 郭旺珍*, 张天真

(南京农业大学作物遗传与种质创新国家重点实验室, 南京 210095)

摘要:通过陆地棉高品质纤维种质系 7235 的棉纤维混合 cDNA 文库随机测序, 得到一个开放阅读框(open reading frame, ORF)完整的棉纤维表达蛋白(GhCFE, GenBank 登录号:DQ073045)cDNA 序列。该 cDNA 序列全长 1274bp, 最大的 ORF 为 996bp, 编码 332 个氨基酸, 其理论上的等电点 $pI=6.14$, 分子量 $M_w=37.7\text{KD}$ 。它的 N-末端存在一个长度为 27 个氨基酸残基的信号肽。RT-PCR 分析表明, 该基因在根、茎中不表达, 在叶片中微弱表达, 在不同发育时期的纤维细胞中均表达, 且在纤维伸长期表达量最高。进化树显示 GhCFE 与已报道的 3 个棉纤维表达蛋白 cDNA 有很高的同源性。通过构建酵母表达载体, 发现该基因对酵母细胞的伸长和细胞壁加厚没有显著影响。利用 pBI121 质粒构建了含 35S 启动子的正义表达载体和含棉纤维特异表达启动子 E6 的正义和反义表达载体, 正在通过农杆菌介导的遗传转化, 开展该基因的功能验证。

关键词:棉花; 纤维表达蛋白; 表达载体

中图分类号:S562.035.03 **文献标识码:**A

文章编号:1002-7807(2006)02-0067-07

Cloning and Characterization of A Novel Cotton Fiber Expressed Protein (Gh-CFE)cDNA

GUO Ying, GUO Wang-zhen*, ZHANG Tian-zhen

(State Key Laboratory of Crop Genetics & Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, China)

Abstract: Through randomly sequencing the cotton fiber cDNA library from line 7235 with elite fiber quality in *Gossypium hirsutum* L., a cDNA clone encoding cotton fiber expressed protein (designated as GhCFE, GenBank accession number: DQ073045) was isolated. The full length of this cDNA clone was 1274bp, and its largest open reading frame encoded 332 amino acids. The putative protein of this gene had an isoelectric point of 6.14 and a calculated molecular weight of 37.7KD. It had a signal peptide with 27 amino acid residues in N-terminal. RT-PCR analysis indicated this gene did not express in root or stem but only had a weak expression in leaf, while it could be detected in ovules and fiber cells of different developing periods, especially in elongating fibers. Phylogenetic tree showed GhCFE had high homologies with other CFEs reported before. We cloned GhCFE into the fission yeast (*S. pombe*) vector pREP5N and detected that the gene had no significant effect on elongating cells or thickening cell wall in the transformed yeast. Further, Sense expression vector containing 35S promoter and both sense and antisense expression vectors containing E6 promoter using pBI121 plasmid were constructed, and the work of transferring these recombinant vectors into *Gossypium hirsutum* L.

收稿日期:2005-07-21 作者简介:郭 婕(1981-),女,在读硕士;* 通讯作者,moelab@njau.edu.cn

基金项目:国家自然科学基金(30471104; 30270806), 国家 973 项目(2002CB111301), 教育部新世纪优秀人才项目(NCET-04-0500), 江苏省人才基金(BK2003414)

by *Agrobacterium tumefaciens* is ongoing.

Key words: *Gossypium hirsutum* L.; fiber expressed protein; expression vector

棉纤维是由棉花胚珠外珠被表皮层的单细胞分化发育而成,根据其形成和发育过程已将棉纤维的分化和发育分为四个相互重叠时期:纤维原始细胞分化和突起、初生壁伸长、次生壁增厚和脱水成熟。棉纤维细胞在开花当天开始伸长,纤维伸长是初生壁不断延伸的结果,伸长持续时间取决于遗传和环境因素,一般认为22~28 d。纤维细胞最快的伸长速率发生在开花后6~12 d,在开花后15~20 d,纤维伸长可达最终长度的80%。伴随着次生壁纤维素沉积的加速,开花后20~25 d纤维伸长减弱^[1]。在伸长阶段,纤维细胞长度可为其直径的1000~3000倍(直径为20 μm),纤维最终长度可达20~30 mm,有的可达35~40 mm^[2~3]。

了解棉纤维细胞的表达模式以及可能的生物学功能,系统地阐明棉纤维细胞发育调控的分子机理,是目前棉纤维发育功能基因组研究的重要内容之一。自从首次报道运用cDNA文库差异筛选方法克隆出纤维特异表达基因E6^[4],现已分离出20余个纤维细胞特异表达或优势表达类型的基因。它们大部分在纤维伸长阶段表达,少数在纤维细胞壁增厚期或整个发育过程均能表达^[5]。

棉花的再生系统周期长是目前棉花目的基因功能验证的瓶颈^[6~7]。裂殖酵母(*S. pombe*)是单细胞真核生物,其遗传接近高等真核生物。Xia等^[8]利用酵母系统分离到多个拟南芥的细胞骨架和极性生长相关基因。Ji等^[9]也利用酵母表达系统,将棉纤维中特异表达的Tubulin基因在酵母细胞中过量表达,对其相关功能起到了很好的预示作用。

Yamamoto^[10]利用cDNA差异显示法,从陆地棉Acala SJ-1品系中获得3个同源性高,在棉纤维细胞中优势表达的cDNA克隆,这3个克隆分别被定名为棉纤维表达蛋白(cotton fiber expressed protein,CFE)CFE1、CFE2和CFE3。本文通过对陆地棉优质品系7235的不同纤维发育时期的混合cDNA文库随机测序,筛选获得了一个新的CFE cDNA克隆,并对其进行了表达特征及功能的初步分析。

1 材料和方法

1.1 材料

1.1.1 植物材料。供试材料为陆地棉高品质纤维种质系7235,种植于南京农业大学网室。
1.1.2 菌种、质粒及试剂。植物表达载体pBI121-CaMV35S启动子、pBI121-E6启动子和*E. coli*菌株DH5α均为本实验室保存。裂殖酵母(*Schizosaccharomyces pombe*)菌株S.P.Q01和*S. pombe/E. coli*穿梭载体pREP5N由中科院微生物研究所夏桂先提供。pGEM T-Vector和T4连接酶购自Promega公司,限制性内切酶为MBI产品。日常型质粒DNA小量快速制备试剂盒(Rapid Plasmid DNA Daily Mini-prep kit)为V-GENE BIOTECHNOLOGY LIMITED产品,小量胶回收试剂盒(Gel Extraction Mini Kit)为上海华舜生物工程有限公司产品。

1.2 方法

1.2.1 unigene获得及序列分析。通过构建陆地棉优质材料7235开花后5、8、11、14、17、20、23、25 d等不同纤维发育时期混合cDNA文库,随机挑取6000余个克隆进行5'端测序分析,获1056个非冗余contig并对其功能初步预测。根据功能分类,选取与纤维表达蛋白相关的cDNA克隆,对插入片段进行全序列测定。利用DNAMAN软件完成ORF预测,在Genebank数据库进行核酸和蛋白的Blast比对,采用SignalP程序进行信号肽预测,序列位点分析采用Motif Scan程序,进化树采用DNAMAN软件的Alignment程序进行。构建cDNA文库的克隆载体为pBluescript II SK(+),随机克隆测序由上海联合基因公司完成。

1.2.2 总RNA提取。采收萌发10 d后的棉花幼苗的根、茎、植株嫩叶;0、1、3DPA的胚珠和纤维混合物,5、8、11、14、17、20、23DPA的纤维,采用本实验室发展的CTAB-酸酚法^[11~12]提取陆地棉7235的根、茎、叶和不同纤维发育时期的纤维细胞总RNA。

1.2.3 RT-PCR。取1 μg经DNase I消化后的总RNA,利用RevertAid™ First Strand cDNA Synthesis Kit(MBI)进行cDNA第一链合成,然后取0.5 μL反转录产物进行RT-PCR,用于扩增

的引物分别为 5' TGT, GGC, GGT, TAA, AGA, TG 3' 和 5' GAC, TCA, GCG, ACG, GTT, C 3', 扩增条件为 95°C 预变性 5 min; 94°C 45 s, 51°C 45 s, 72°C 1 min, 30 个循环; 72°C 延伸 10 min。同时为了校正 PCR 模板的相对浓度, 用组成性表达的 EF1 α 一对特异引物, 作为内标进行平行 PCR 扩增。EF1 α 的特异引物为 F: 5' AGA, CCA, CCA, AGT, ACT, ACT, GCA, C 3' 和 R: 5' CCA, CCA, ATC, TTG, TAC, ACA, TCC 3'。

1.2.4 酵母表达载体的构建及转化鉴定。

pREP5N 载体携带 Leu2 基因作为选择标记, 并且还带有一个 VB1 抑制启动子 nmt-1, 它可以调控插入多克隆位点的目标基因的表达。目的基因两侧有原克隆载体 pBluescript II SK(+) 的 XhoI 和 NotI 酶切位点, 酵母表达载体 pREP5N 的多克隆位点也有这两个酶切位点, 且方向相同。用 XhoI 和 NotI 双酶切 pBluescript II SK(+) (带有目的基因) 和 pREP5N, 将目的片段回收连接(图 1)。酵母空载体和转化体的长, 宽及长/宽的显著性分析采用 t 测验统计法。

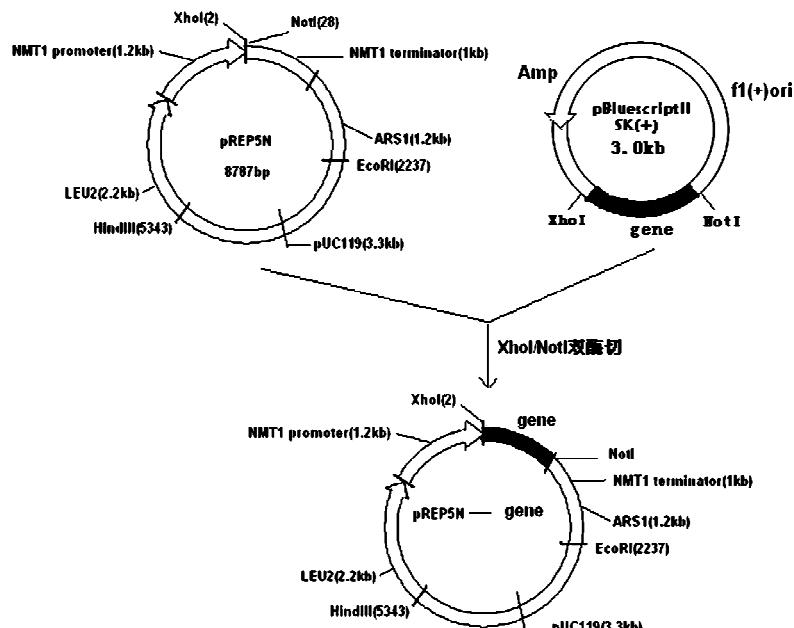


图 1 酵母表达载体的构建

Fig. 1 Construction of the fission yeast expression vector

将构建好的酵母表达载体 pREP5N-GhCFE 通过电击转化的方法转化酵母细胞, 转化株在含有 75 mg·L⁻¹ adenine 和 uracil 的 minimal medium (MM) 培养基平板上筛选, 将 VB1 加入培养基至终浓度为 2 μ mol·L⁻¹ 以抑制 nmt-1 启动子。菌落首先在含有 VB1 的液体 MM 培养基中生长至指数生长阶段, 用不含 VB1 的 MM 培养基洗三次来解除对启动子的抑制作用。然后继续培养 22 个小时。将细胞用 DAPI 染色, 在荧光显微镜下观测。

2 结果与讨论

2.1 与纤维表达蛋白相关 cDNA 序列分析

通过对从陆地棉优质材料 7235 棉纤维发育混合 cDNA 文库中获得的 1056 个 uni-EST 序列进行功能预测, 获得 3 个与纤维表达蛋白相关的克隆。分别对其插入片段全序列测序并进行序列

分析, 获得一个 ORF 完整的 cDNA 克隆。该 cDNA 克隆长度为 1274 bp, 最大的 ORF 为 996 bp, 编码 332 个氨基酸(图 2), 被定名为 GhCFE (GenBank 登录号: DQ073045)。用 Expasy pI/Mw 程序对其氨基酸序列进行了一级结构的预测, 其理论上的等电点 pI = 6.14, 分子量 Mw = 37.7 KD。Blastn 结果显示, 该 cDNA 序列与已报道的 3 个陆地棉纤维表达蛋白 cDNA 克隆相似性均很高, 与 CFE3, CFE2 和 CFE1 分别达到了 98%, 97% 和 95%。与苜蓿(*Medicago truncatula*), 拟南芥(*Arabidopsis thaliana*) I 号染色体上的一个基因也都有一定的相似性。SignalP 运行程序预测到该序列的 N-末端存在一个长度为 27 个氨基酸残基的信号肽。说明该 cDNA 序列编码的是一胞外分泌蛋白。Motif Scan 程序表明, 这个序列的蛋白激酶 C 磷酸化位点位于 10~12, 72~74, 185~187, 218~220, 263~265, 283~

285; N-豆蔻酰化位点在 22~27, 26~31, 64~69; 酪蛋白 II 磷酸化位点在 108~111, 136~139, 156~159, 164~167, 263~266, 293~296; 酪氨酸硫酸盐位点在 115~129。

为进一步明确克隆到的这个基因与 Yamamoto^[9] 报道的三个陆地棉纤维表达蛋白基

因的进化关系,将这四个 cDNA 的 ORF 所翻译的氨基酸进行了多重序列比较。结果表明,这个基因与 CFE2 属于同一分支(图 3)。并且这四个基因都具有相同的 4 个蛋白激酶 C 磷酸化位点,2 个 N-豆蔻酰化位点,3 个酪蛋白 II 磷酸化位点以及酪氨酸硫酸盐位点(图 4)。

* 表明终止密码子位置

图 2 GhCFE 的全长序列及推导的氨基酸序列

Fig. 2 The full length nucleotide sequence of GhCFE and its deduced amino acid sequence

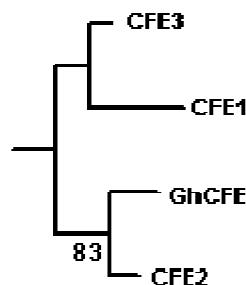


图 3 GhCFE 与已经报道的 CFE1、CFE2、CFE3 比较的进化树 (分支结上示自展值)

Fig. 3 Phylogenetic-Tree of GhCFE compared with other CFEs reported

GhCFE	MATASTWILSLKVFLISTGILGIVLGLKISVPLVLEFSVS	40
CFE2	MATASTWILSLKVFLISTGILGIVLGLKISVPLVLEFSVS	40
CFE3	MATASTWILSLKVFLISTGILGIVLGLKISVPLVLEFSVS	40
CFE1	MASASTWILSLKILLISTGILGIALGLKISVPLVLEFSVS	40
Consensus	ma astwilslk listgilgi lglkisvplvlefsvs	
GhCFE	QAPLWWSGFRSWLKPPYLYVVINGIITIAASSRFNQNNG	80
CFE2	QAPLWWSGFRSWLKPPYLYVVINGIITIAASSRFNQNNG	80
CFE3	QAPLWWSGFRSWLKPPYLYVVINGIITIAASSRFNQNNG	80
CFE1	QAPLWWSGFRSWLKPPYLYVVINGIITIAASSRFNQNNG	80
Consensus	qaplwwsgfrswlkppylvvvingiitiaassrfnqnng	
GhCFE	EKDQMEQMOPRKPISADQQPMVEYETKSGWDSDAVESSDF	120
CFE2	EKDQMEQMOPRKPISADQQPMVEYETKSGWDSDAVESSDF	120
CFE3	EKDQMEQMOPRKPISEDQQPIVEYDTKSGWGSDAVESSDF	120
CFE1	EKDQMEQMOPRKPISEDQQPIVEYDTKSGWGSDAVESSDF	120
Consensus	ekdqmeqmoprpkis dqgp vey tksgw sdavessdf	
GhCFE	VYEENQRGEEVETRVSEEESNVAVKDDRDGNEFVISKSEW	160
CFE2	VYEENQRGEEVETRVSEEESNVAVEDDRDGNEFVISKSEW	160
CFE3	VYEENQRGEEVATRVSEEESNVAVEDDRDGNEFVISKSEW	160
CFE1	VYEENQRGEEVATRVSEEESNVAVEDDRDGNEFVISKSEW	160
Consensus	vyeenqrgeev trvseeesnvav ddrgn nefvisksew	
GhCFE	IPPSRTDSSEIPLDALLIQEKPAPSSRFGHRKPVKANPEG	200
CFE2	IPPSRTDSSEIPLDALLIQEKPAPSSRFGHRKPVKVNPEG	200
CFE3	IPPSRTDSSEIPLDALLIQEKPAPSSRFGHRKPVKVNPEG	200
CFE1	IPPSRTDSSEIPLDALLIQEKPAPSSRGHRKPVKVNPEG	200
Consensus	ippsrtdsseipldalliqekpapssr ghrkpvk nppeg	
GhCFE	GRALKAAKPKRRETLENTWKMITEGKSMPLSRHLKKSDTW	240
CFE2	GRALKVAKPKRRETLENTWKMITEGKSMPLSRHLKKSDTW	240
CFE3	GRALKAAKPKRRETLENTWKMITEGKSMP	229
CFE1	GRALKVAKPKRRETLENTWKMITEGKSMPLSRHLKKSDTW	240
Consensus	gralk apkkrhetlentwkmitegksmp	
GhCFE	ENHGRDINMEALTSSPLMKKSETFRDRTNYQLPPEQVSSF	280
CFE2	ENHGRDINVEALTSSPLMKKSETFRDRTNYQLPPE	275
CFE3		
CFE1	ENHGRDINVEALTSSPLMKKSETFRDRTNYQLPPEQVSSF	280
Consensus		
GhCFE	PASGKLRLKEPSLSQDELNRVVEAFIKKFNDEMRLQRQESL	320
CFE2		
CFE3		
CFE1	PASGKLRLKEPSLSQDELNRVVEAFIKKFNDEMRLQRQESL	320
Consensus		
GhCFE	NQYMEMVDRGS	331
CFE2		
CFE3		
CFE1	NQYMEMVNRGS	331
Consensus		

图 4 GhCFE 与已经报道的 CFE1、CFE2、CFE3 的同源性比较

Fig. 4 Alignment of GhCFE with CFEs reported

2.2 GhCFE 基因的表达分析

为研究 GhCFE 在棉花中表达的时空特点, 对根、茎、叶和不同时期棉纤维组织来源的 RNA 进行 RT-PCR 分析。结果表明, 该基因在根、茎中都不表达, 在叶中微弱表达, 在胚珠和纤维细胞中表达, 并在开花后 5~20 d 的纤维中持续高效表

达, 20 d 后减弱(图 5)。Yamamoto 等^[9]在分析 CFE1, CFE2 和 CFE3 时也发现这三个基因在棉纤维中表达量很高, 特别是在开花后 4 d 完整的胚珠中就可检测到 CFE1 的转录, 在开花后 10 d CFE1 转录达到高峰, 15 d 后开始减弱。



1: 根; 2: 茎; 3: 叶; 4-6: 开花后 0, 1, 3 天胚珠; 7-13: 分别为开花后 5, 8, 11, 14, 17, 20, 23 d 的纤维细胞; EF1 α 为内标

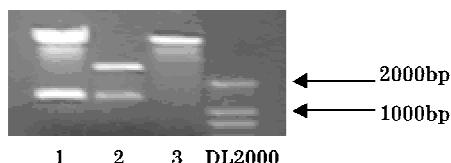
图 5 GhCFE 基因的时空表达特征

Fig. 5 The expressed patterns of CFE in *Gossypium hirsutum* L.

2.3 酵母表达载体的构建及转化鉴定

根据 RT-PCR 分析结果, 推测 GhCFE 是在纤维伸长阶段优势表达的一个基因。为了明确其是否与棉纤维细胞的极性生长有关, 进一步构建了 GhCFE 的酵母表达载体, 并进行 *E. coli* 菌株 DH5 α 转化研究。将转化后筛选到的阳性克隆在 LB 培养基(加氨苄青霉素, 50 mg·L⁻¹)中培养过

夜后, 提质粒, 进行表达载体的酶切验证。*Xba*I/*Not*I 分别双酶切 pREP5N、pBluescript II SK (+)-GhCFE 和 pREP5N-GhCFE, 构建好的酵母表达载体酶切后得到的小片段大小符合目的基因大小 1300bp 左右, 大片段大小与双酶切后的 pREP5N 大小相同。证明所构建载体正确(图 6)。



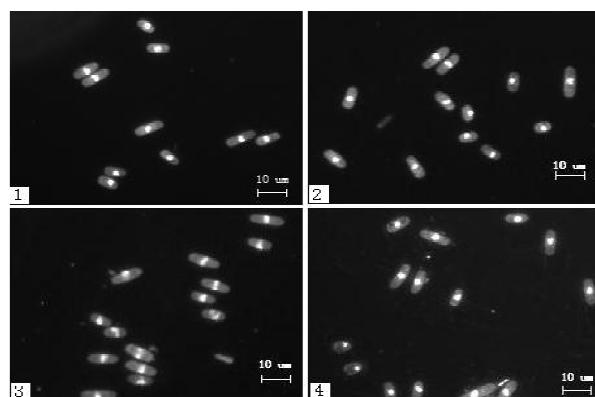
1: *Xba*I/*Not*I 双酶切 pREP5N-GhCFE; 2: *Xba*I/*Not*I 双酶切 pBluescript II SK (+)-GhCFE; 3: *Xba*I/*Not*I 双酶切 pREP5N

图 6 酵母表达载体的酶切分析

Fig. 6 Plasmid restriction analysis(*Xba*I/*Not*I)

将构建好的酵母表达载体 pREP5N-GhCFE 通过电击转化法转化酵母细胞, 转化体形态学观察表明, 生长在诱导培养基中且带有 pREP5N-GhCFE 载体的细胞与生长在同样培养基带有

pREP5N 的细胞在长度, 宽度及长/宽上并没有显著差异(图 7)。说明该基因的表达与细胞的极性生长无关。



1: pREP5N, 诱导; 2: pREP5N, 未诱导; 3: pREP5N-GhCFE, 诱导; 4: pREP5N-GhCFE, 未诱导

图 7 GhCFE 在酵母细胞中的表达

Fig. 7 Over expression of GhCFE in fission yeast cells

为了进一步阐明 GhCFE 基因在棉纤维发育中的可能功能以及对棉纤维品质的影响, 分别对该基因构建了含 pBI121-CaMV35S 启动子的正义表达载体和含 pBI121-E6 启动子的正义和反义表达载体, PCR 检测和酶切验证表明所构建载体正确。该表达载体正通过农杆菌介导的遗传转化法进行转基因棉花的功能验证。

尽管对于这个基因的详细作用机理及生化途径还没有完全了解,但是,酵母表达的结果表明该基因与细胞的极性生长无关。来自棉花不同组织和不同棉纤维发育时期的 RT-PCR 分析,表明该基因是一个在棉纤维伸长阶段优势表达的基因,应该在棉纤维伸长过程中有一定作用。同时,棉花纤维表达蛋白基因目前已报道了 4 个,序列分析表明它们是一类同源性非常高的基因,属于一个基因家族,Yamamoto 推测这类基因可能位于四倍体陆地棉的某一部分同源染色体上。因此,下一步的研究工作我们将集中在 4 个 CFE 基因的染色体分布特点研究;利用已构建好的植物正义、反义表达载体,通过转基因技术进行 GhCFE 基因可能作用机理的探讨。

致谢:

本研究所用的裂殖酵母菌株 S. P Q01 和穿梭载体 pREP5N 由中科院微生物研究所夏桂先先生提供。在研究方法上,也得到夏桂先先生及课题组研究人员的指导,特此致谢。

参考文献:

- [1] 刘继华,杨洪博,曹鸿鸣. 棉纤维生长发育(3)棉花纤维的伸长发育[J]. 中国棉花,1995,22(4):38-39.

- [2] 杜雄明,潘家驹,汪若海. 棉纤维细胞分化和发育[J]. 棉花学报,2000,12(4):212-217.
- [3] 孙 杰,李艳甲,李园莉,等. 棉花纤维特异表达基因 GhF1 的分离及鉴定[J]. 棉花学报,2005,17(5):259-263.
- [4] JOHN M E, Crow L J. Gene expression in cotton (*Gossypium hirsutum* L.) fiber: cloning of the mRNAs[J]. Proc Natl Acad Sci USA, 1992, 89: 5769-5773.
- [5] 郭旺珍,孙 敬,张天真. 棉花纤维品质基因的克隆与分子育种[J]. 科学通报,2003,48(5):411-417.
- [6] WILKINS T, Rajasekaran K, Anderson D M. Cotton biotechnology[J]. Crit Rev Plant Sci, 2000, 19: 511-550.
- [7] 张天真. 棉花纤维品质分子育种的现状及展望[J]. 棉花学报,2000,12(6):321-326.
- [8] XIA G X, Ramachandran S, Hong Y, et al. Identification of plant cytoskeletal, cell cycle-related and polarity-related proteins using *Schizosaccharomyces pombe*[J]. The Plant Journal, 1996, 10(4):761-769.
- [9] JI S J, Lu Y C, Li J, et al. A β -tubulin-like cDNA expressed specifically in elongating cotton fibers induces longitudinal growth of fission yeast[J]. Biochemical and Biophysical Research Communications, 2002, 296:1245-1250.
- [10] YAMAMOTO E, Baird W V. Three Cotton Fiber-Expressed cDNAs (Accession Nos. AF072404, AF072405, and AF072406)[J]. (PGR98-144) Plant Physiol, 117:1525.
- [11] 将建雄,张天真. 利用 CTAB 酸酚法提取棉花组织总 RNA[J]. 棉花学报, 2003, 15(3):166-167.
- [12] 刘 康,胡凤萍,张天真. 棉花胚珠与纤维蛋白的两种提取方法比较研究[J]. 棉花学报, 2005, 17(6): 323-327. ●