Original: *Cotton Science*, 2006,18(2):67~73

# Cloning and Characterization of a Novel Cotton Fiber Expressed Protein (*GhCFE*) 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 7235 germplasm line with elite fiber quality in *Gossypium hirsutum* L., a cDNA clone encoding cotton fiber expressed protein was isolated, designated as *GhCFE* (GenBank accession number: DQ073045). The full length of this cDNA clone was 1274 bp, with the 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 37.7 kD. 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, only had a weak expression in leaf, while it could be detected in ovules and fiber cells of different developmental stages, especially in elongating fibers. Phylogenetic tree showed *GhCFE* had high homologies with other CFEs reported previously. We cloned *GhCFE* into the fission yeast (*S. pombe*) vector pREP5N and detected that this 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 recombined vectors into *Gossypium hirsutum* L. by *Agrobacterium tumefaciens* is ongoing.

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

Cotton (*Gossypium* spp.) is an important cash crop and the second largest sources of textile fiber and edible oil throughout the world. Cotton fibers used in textiles are derived from epidermal cells from outer integuments of ovules. The growth of cotton fiber is characterized by four major distinct yet overlapping developmental stages: initiation (-3 ~ 0 DPA, day post anthesis), cell elongation and primary wall synthesis (1~25 DPA), cell wall deposition (16 ~40 DPA) and maturation (40~50 DPA). Cotton fiber initiation is a quasi-synchronous process in developing ovules during anthesis. The signals important to fiber cell differentiation must occur prior to the formation of fiber cell initials. Fiber initials first become visible at the day of flowering (0 DPA), although the physiological changes arise before this day. The complex synthesis of primary wall commences at the first two stages of the fiber development, and goes on to 25 DPA. With the acceleration of the cellulose deposition, the fiber elongation becomes tardiness and extends to the 80% length of final fiber length at 20~25 DPA[1]. In fiber elongation stage, the length of fiber cell can extend to 20~30 mm, which is 1000~3000 fold than the diameter[2-3].

Today, the hot spot of cotton fiber research is functional genomics, including the expression pattern and putative biological function, the mo-

**Received**: 2005-07-21 **Author**: Guo Ying (1981-), *: Corresponding author, moelab@njau.edu.cn

**Sponsor**: National Natural Science Foundation of China (30471104; 30270806), the State Key Basic Research and Development Plan of China (2002CB111301), the Program for New Century Excellent Talents in University (NCET-04-0500), and Science Foundation in Jiangsu Province (BK2003414).

lecular foundation of developmental regulation for fiber cells. More than twenty structural genes expressed in elongated fiber have been isolated since John and Crow cloned E6 from *G. hirsutum*[4]. Most of them were detected with the highest expression level in elongated fibers, a few of them were expressed in the deposition or the whole stages[5].

One bottleneck of the functional analyses is the long period for the cotton regeneration[6-7]. The fission yeast (*Schizosaccharomyces pombe*) is a single cell eukaryote, characterized by the similar genetic pattern to higher eukaryotes. In *Arabidopsis*, the genes of cell framework and polar-growth were isolated by *S. pombe* phenotypic screening[8]. A β-tubulin-like cDNA, specially expressed in cotton fibers was predicted the function with the excessive expression in yeast cells[9].

In this study, three cotton fiber-expressed genes, named *CFE*1, *CFE*2, and *CFE*3, were isolated by the method of cDNA differential display, from upland cotton variety. Acala SJ-1[10]. A cDNA clone homologue to *CFE* was isolated from the cDNA library of cotton fiber, followed by the functional analysis.

## 1 Materials and Methods

### 1.1 Materials

**1.1.1** *Plant materials.* Upland cotton line 7235 with elite fiber quality was grown under standard field conditions at Nanjing Agricultural University.

**1.1.2** *Bacteria strains, plasmid and reagents.* Expression vector pBI121-CaMV35S, pBI121-E6 and *Escherichia* coli DH5a were preserved by Nanjing Agricultural University. *Schizosaccharomyces pombe* var. S. P Q01 and pREP5N vector were provided by Prof. Xia Gui-xian from the Chinese Academy of Sciences. pGEM T-Vector and T4 ligase were selected for ligation (Promega, Madison, USA). Restriction endonuclease was purchased from MBI (Fermentas, Lithuania). Plasmid extraction kit was purchased from V-gene Biotechnology Ltd (Zhejiang, China). Gel Extraction Mini Kit was purchased from Watson Biotechnologies, Inc (Shanghai, China).

### 1.2 Methods

**1.2.1** *Acquisition of the unigenes and sequence analysis.* A mixed development fiber cDNA library was constructed containing 7235 fibers at 5, 8, 11, 14, 17, 20, 23 and 25 DPA (days post anthesis). In total, 1056 non-redundant clones were obtained by 5′ sequence analysis by randomly sequencing more than 6000 clones. According to functional categories, a few cDNAs related to cotton fiber expression genes were selected to conduct further full-length sequencing. DNA sequence analysis, ORF searches and alignment analysis were carried out by DNAMAN software. Nucleic acid and protein sequence BLAST were based on GenBank database (http://www.ncbi.nlm.nih.gov/blast). Signal peptide and sequence motif analysis were predicted by SignalP (www.cbs.dtu.dk/services/SignalP/) and Motif scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan), respectively. The cloning vector used to construct cDNA library is pBluescript II SK (+) (Merck, Darmstadt, Germany). Automated DNA sequencing was completed in United Gene of Shanghai, Ltd (Shanghai, China).

**1.2.2** *RNA extraction.* Total RNAs were extracted from upland cotton 7235 seedling roots, stems and leaves germinated for 10 days, 0, 1, 3 DPA ovules and fibers, 5, 8, 11, 14, 17, 20, 23 DPA fibers by the CTAB-acidicn phenolic method[11-12].

**1.2.3** *RT-PCR.* First-strand cDNA was synthesized from 1 μg total RNA (DNase I digested) using RevertAidTM First Strand cDNA Synthesis Kit (Fermentas, Lithuania). Aliquots (0.5 μL) of the RT products were then amplified using the primes F: 5′ TGTGGCGGTTAAAGATG 3′ and R: 5′ GACTCAGCGACGGTTC 3′. The reaction was denatured at 95℃ for 5 min, followed by 30 cycles of 94℃ for 45 s,

51℃ for 45 s, and 72℃ for 1 min, incubated at 72℃ for 10 min as the last cycle. The gene *eF*1$\alpha$ was used as an internal control. Primers for *eF*1$\alpha$ were described as F: 5′-AGACCACCAAG-TACTACTGCAC-3′, R: 5′-CCACCAATCTT-GTACACATCC -3′.

**1.2.4** *Construction of yeast expression vector and transformation.* pREP5N vector with a repressive promoter, nmt-1, controls the expression of the genes in multi-clone region (Invitrogen, Carlsbad, CA). This vector contains a selective marker, *Leu*2. Two enzyme sites for objective genes of pBluescript II SK (+) vector, *XhoI* and *NotI*, had the same direction with pREP5N vector. The objective nucleotide fragment was cloned into pREP5N vector (Fig. 1). T test was employed for the analysis of the length, width and length/with for yeast vectors.

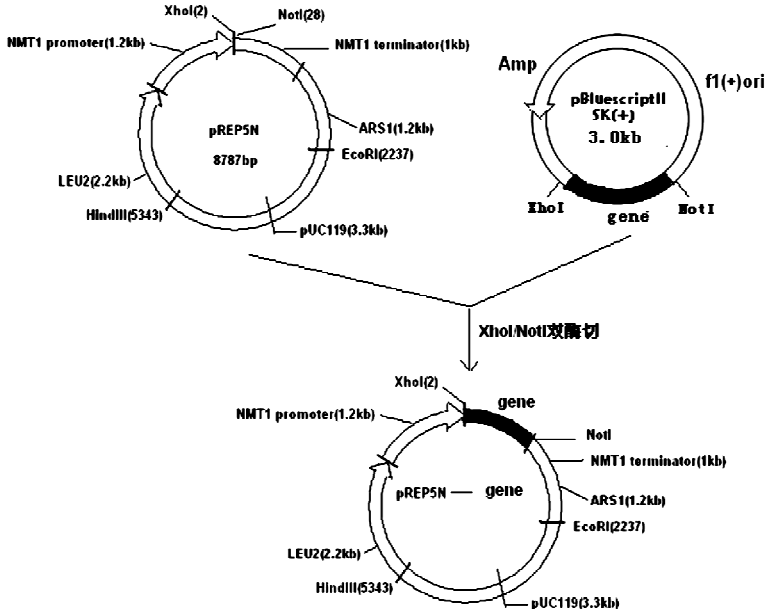



**Fig. 1 Construction of the fission yeast expression vector**

The pREP5N-GhCFE vector was transferred into yeast cells by electric method. The yeast cells were incubated on the minimal medium (MM) plat contained adenine and uracil. In order to repress the nmt-1 promoter, VB1 was added into the medium with a final content of 2 $\mu$mol • $L^{-1}$. The yeast colonies were grown to the logarithm growth stage on the medium with VB1. Then, the yeast cells were incubated for 22 h after rinsed by medium without VB1 for three times. The fluorescent microscope was used to observe the yeast cells dyeing by DAPI.

## 2 Results and Analysis

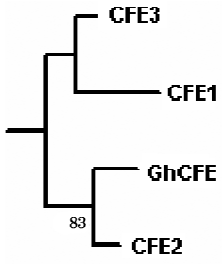
### 2.1 Identification and sequence analysis of fiber expressed protein cDNA

Three clones with identity with fiber expressed protein were acquired by functional prediction analysis of 1056 unigene sequences from mixed cDNA library. The cDNA contained the whole insertion of 1274 bp with the complete ORF of 996 bp, encoding 332 amino acids (Fig. 2). This gene was named *GhCFE* (GenBank Acc No DQ073045). The putative protein was predicated with software Expasy PI / MW to have an isoelectric point of 6.14 and a calculated molecular weight of 37.7 kD (www.expasy.ch/tools/pi_tool.html). The result of BLASTn indicated that this gene shared 98%, 97%, and 95% identity with three upland cotton fiber expressed protein cDNA clones, *CFE*3, *CFE*2, and *CFE*1, respectively. *GhCFE* shared identity with a gene in Medicago truncatula (Chr. I). *GhCFE* was predicted with signal peptide′s structure of the N-terminal in amino acids, contained 27 amino acids, by SignalP software. The result of Motif Scan software showed that there were protein kinase C phosphorylated sites in

```
1       gaaaccacgctaacccaaaacacaagaaaaagaaagaaagaaaaaaaaaaaaggaatacc
61      ccttaccttcatagagagaacaaaagaaacaaagcaaggtatttcttttcttttcttcac
121     ctgtaaaatggcgacggcgagtacttggatattgtcgctaaaggtatttttaatttctac
              M  A  T  A  S  T  W  I  L  S  L  K  V  F  L  I  S  T
181     cggtatattgggtatagttttaggacttaaaatctctgttccattggttttggaattctc
         G  I  L  G  I  V  L  G  L  K  I  S  V  P  L  V  L  E  F  S
241     tgtttctcaagctccgttatggtggagtggtttccgttcttggctcaagcctccatatct
         V  S  Q  A  P  L  W  W  S  G  F  R  S  W  L  K  P  P  Y  L
301     ttacgtcgtcatcaacggaatcatcatcacaatagcagcatcgtcgcggtttaaccaaaa
         Y  V  V  I  N  G  I  I  I  T  I  A  A  S  S  R  F  N  Q  N
361     caacggcgagaaagatcagatggagcagatgcaaccgcggcctaagatctcggcggatca
         N  G  E  K  D  Q  M  E  Q  M  Q  P  R  P  K  I  S  A  D  Q
421     acaaccaatggtggagtatgagacaaagagcgggtgggactctgatgcagtggaatccag
         Q  P  M  V  E  Y  E  T  K  S  G  W  D  S  D  A  V  E  S  S
481     tgatttcgtgtacgaggaaaatcagagaggagaagaggtggaaaccagggtttccgagga
         D  F  V  Y  E  E  N  Q  R  G  E  E  V  E  T  R  V  S  E  E
541     ggagagcaatgtggcggttaaagatgacagagatggaaacgagtttgttatctctaagtc
         E  S  N  V  A  V  K  D  D  R  D  G  N  E  F  V  I  S  K  S
601     ggagtggattcctccaagtagaacggattcctcggagattccgttggatgctctacttat
         E  W  I  P  P  S  R  T  D  S  S  E  I  P  L  D  A  L  L  I
661     acaggagaaacctgcgccttcttctagattcggtcaccggaaacctgttaaagccaatcc
         Q  E  K  P  A  P  S  S  R  F  G  H  R  K  P  V  K  A  N  P
721     cgaaggtgggcgagcgttgaaagcggcgaagccaaaacggcatgagacgctggaaaacac
         E  G  G  R  A  L  K  A  A  K  P  K  R  H  E  T  L  E  N  T
781     ttggaaaatgataacggaagggaaatcaatgccgttgtccagacacttgaagaagtcaga
         W  K  M  I  T  E  G  K  S  M  P  L  S  R  H  L  K  K  S  D
841     cacgtgggagaatcacggccgtgatatcaacatggaggcgttgaccagctccccctctgat
         T  W  E  N  H  G  R  D  I  N  M  E  A  L  T  S  S  P  L  M
901     gaaaaaatcggaaacgttcagagaccggaccaattaccagctgccacccgaacaagtaag
         K  K  S  E  T  F  R  D  R  T  N  Y  Q  L  P  P  E  Q  V  S
961     ctctttcccggcttcaggaaagctgagaaaagaaccgtcgctgagtcaggacgagttaaa
         S  F  P  A  S  G  K  L  R  K  E  P  S  L  S  Q  D  E  L  N
1021    tcgtcgagtggaagctttataaagaagtttaatgacgagatgaggttacagagacaaga
         R  R  V  E  A  F  I  K  K  F  N  D  E  M  R  L  Q  R  Q  E
1081    atcacttaatcagtacatggaaatggttgaccgtggaagttagcccctaaacacaataga
         S  L  N  Q  Y  M  E  M  V  D  R  G  S  *
1141    ccaatatcttctgccataaccctcaagagaagaaaaaaatatatattagaataatttttg
1201    ggaatatgctttgctccaataaatttaaaaaggttggagatttgaaaaaaaaaaaaaaaa
1261    aaaaaaaaaaaaa
```

* indicates the positions of termination codons

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




Note: Bootstrap values (%) based on 1000 replicates are indicated beside nodes.

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

```
GhCFE       MATASTWILSLKVFLISTGILGIVLGLKISVPLVLEFSVS   40
CFE2        MATASTWILSLKVFLISTGILGIVLGLKISVPLVLEFSVS   40
CFE3        MATASTWILSLKVFLISTGILGIVLGLKISVPLVLEFSVS   40
CFE1        MASASTWILSLKLLLISTGILGIALGLKISVPLVLEFSVS   40
Consensus   ma astwilslk  listgilgi lglkisvplvlefsvs
GhCFE       QAPLWWSGFRSWLKPPYLYVVINGIIITIAASSRFNQNNG   80
CFE2        QAPLWWSGFRSWLKPPYLYVVINGIIITIAASSRFNQNNG   80
CFE3        QAPLWWSGFRSWLKPPYLYVVINGIIITIAASSRFNQNNG   80
CFE1        QAPLWWSGFRSWLKPPYLYVVINGIIITIAASSRFNQNNG   80
Consensus   qaplwwsgfrswlkppylyvvingiiitiaassrfnqnng
GhCFE       EKDQMEQMQPRPKISADQQPMVEYETKSGWDSDAVESSDF  120
CFE2        EKDQMEQMQPRPKISADQQPMVEYETKSGWDSDAVESSDF  120
CFE3        EKDQMEQMQPRPKISEDQQPIVEYDTKSGWGSDAVESSDF  120
CFE1        EKDQMEQMQPRPKISEDQQPIVEYDTKSGWGSDAVESSDF  120
Consensus   ekdqmeqmqprpkis dqqp vey tksgw sdavessdf
GhCFE       VYEENQRGEEVETRVSEEESNVAVKDDRDGNEFVISKSEW  160
CFE2        VYEENQRGEEVETRVSEEESNVAVEDDRDGNEFVISKSEW  160
CFE3        VYEENQRGEEVATRVSEEESNVAVEDDRDGNEFVISKSEW  160
CFE1        VYEENQRGEEVATRVSEEESNVAVEDDRDGNEFVISKSEW  160
Consensus   vyeenqrgeev trvseeesnvav ddrdgnefvisksew
GhCFE       IPPSRTDSSEIPLDALLIQEKPAPSSRFGHRKPVKANPEG  200
CFE2        IPPSRTDSSEIPLDALLIQEKPAPSSRFGHRKPVKVNPEG  200
CFE3        IPPSRTDSSEIPLDALLIQEKPAPSSRFGHRKPVKVNPEG  200
CFE1        IPPSRTDSSEIPLDALLIQEKPAPSSRSGHRKPVKVNPEG  200
Consensus   ippsrtdsseipldalliqekpapssr ghrkpvk npeg
GhCFE       GRALKAAKPKRHETLENTWKMITEGKSMPLSRHLKKSDTW  240
CFE2        GRALKVAKPKRHETLENTWKMITEGKSMPLSRHLKKSDTW  240
CFE3        GRALKAAKPKRHETLENTWKMITEGKSMP             229
CFE1        GRALKAAKPKRHETLENTWKMITEGKSMPLSRHLKKSDTW  240
Consensus   gralk akpkrhetlentwkmitegksmp
GhCFE       ENHGRDINMEALTSSPLMKKSETFRDRTNYQLPPEQVSSF  280
CFE2        ENHGRDINVEALTSSPLMKKSETFRDRTNYQLPPE       275
CFE3
CFE1        ENHGRDINVEALTSSPLMKKSETFRDRTNYQLPPEQVSSF  280
Consensus
GhCFE       PASGKLRKEPSLSQDELNRRVEAFIKKFNDEMRLQRQESL  320
CFE2
CFE3
CFE1        PASGKLRKEPSLSQDELNRRVEAFIKKFNDEMRLQRQESL  320
Consensus
GhCFE       NQYMEMVDRGS                               331
CFE2
CFE3
CFE1        NQYMEMVNRGS                               331
Consensus
```

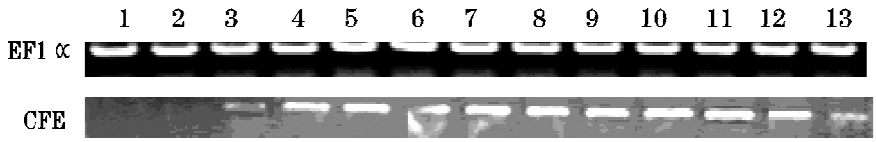
**Fig. 4 Alignment of *GhCFE* with CFEs reported**

10～12, 72～74, 185～187, 218～220, 263～265, 283～285, myristoylation sites in 22～27, 26～31, 64～69, casein phosphorylated sites in 108～111, 136～139, 156～159, 164～167, 263～266, 293～296, and tyrosine vitriolorylated site in 115～129.

In order to illustrate the relation of *GhCFE*, *CFE*1, *CFE*2, and *CFE*3[9], the protein sequences encoded by four CFE genes were analyzed using alignment method. The results showed that *GhCFE* and *CFE*2 located on the same branch of phylogenetic tree (Fig. 3). These four genes had the same sites for protein kinase C phosphorylation, myristoylation, casein phosphorylation, and tyrosine vitriolorylation (Fig. 4).

### 2.2 Expression pattern of *GhCFE*

The spatial and temporal expression pattern of *GhCFE* was detected by RT-PCR analysis in the mRNA from different stages of fiber tissue as well as root, stem, leaf. The results showed that *GhCFE* expressed in ovules and fiber cells, with weak expression level in leaf and no expression level in root and stem. *GhCFE* had high expression level in 5～20 DPA fiber cells, reduced the transcripts after 20 DPA (Fig 5). The results were consistent with previous report that *CFE*1, 2, 3 had high expression level in elongated fiber, with the expression peak in 10 DPA, reduced the transcripts after 15 DPA[9].



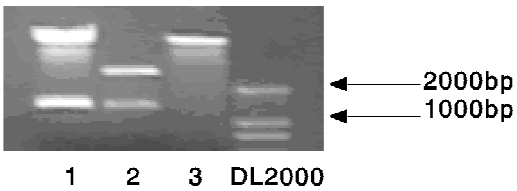

Lane1:Root; Lane2:Stem; Lane3:Leaf; Lane4-6:Ovules of 0, 1 and 3DPA; Lane7-13:Fiber cells of 5, 8, 11, 14, 17, 20, 23 DPA respectively. eF1a is positive control.

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

### 2.3 Construction of yeast expression vector and transformation

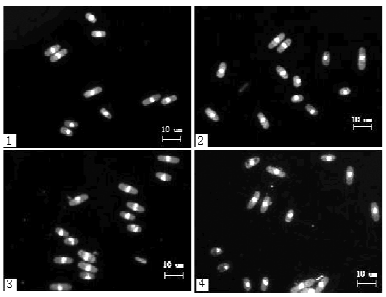
According to the RT-PCR result, *GhCFE* was a gene highly expressed in fiber elongation stage. In order to illustrate the relationship of *GhCFE* and fiber polar growth, yeast expression vector contained *GhCFE* was transferred to E. coli DH5a. The positive clones were incubated in LB medium contained 50 mg · $L^{-1}$ ampicillin overnight. The plasmids were extracted from the positive bacillus and used to digest. pREP5N—*GhCFE* and pBluescript II SK(+)—*GhCFE* were digested by XhoI and NotI. The objective fragment of the yeast vector was 1300 bp, showed the same length with this gene (Fig 6).

The pREP5N-GhCFE vector was transferred into yeast cells by electric method. The observation of fluorescent microscope showed that the yeast cells contained pREP5N-*GhCFE* vector had no significant difference with the control (Fig 7). So, *GhCFE* had not participated in the polar growth of cotton fiber cells.




Lane1:pREP5N—*GhCFE*;
Lane2:pBluescript II SK(+)—*GhCFE*; Lane3:pREP5N

**Fig. 6 Plasmid restriction analysis(XhoI/NotI)**



1: pREP5N, induced; 2: pREP5N, not induced;
3: pREP5N-*GhCFE*, induced; 4: pREP5N-*GhCFE*, not induced

**Fig. 7 Over expression of *GhCFE* in fission yeast cells**

In order to illuminate the biological function of *GhCFE* in cotton fiber development, the sense and anti-sense expression vector were con-

structed into pBI121-CaMV35S and pBI121-E6. These vectors were detected with the correct objective genes after the validation of PCR and digest. The transgenic analysis has been undergoing through the transformation method by *Agrobacterium tumefaciens*.

Although little information was known about its biochemistry path, the function of *GhCFE* was studied primarily. This research showed this gene was not involved in the polar growth of fiber cells. Results from RT-PCR analysis displayed that *GhCFE* had high expression level in fiber elongation stage, which indicated a novel function in fiber elongation. Four fiber expressed protein genes shared the high identity belong to a gene family. Yamamoto suggested that these genes may locate in the same part of homologous chromosomes in allotetraploid cotton. Our further studies will focus on the genetic distribution and transgenic research of these four *CFE* genes.

## Acknowledgments

We thank Prof. Xia Gui-xian (Institute of Microbiology, Chinese Academy of Sciences) for providing the strain S. P Q01 and the protocol.

**References:**

[1] LIU Ji-hua, Yang Hong-bo, Cao Hong-ming. Cotton fiber development (3) fiber elongation development [J]. China Cotton, 1995, 22(4): 38-39 (in Chinese).

[2] DU Xiong-ming, Pan Jia-ju, Wang Ruo-hai. Differentiation and development of cotton fiber cells[J]. Cotton Science, 2000, 12(1): 212-217 (in Chinese).

[3] SUN Jie, Li Yan-jia, Li Yuan-li, et al. Isolation and identification of specifically expressed gene *GhF*1 in cotton fiber cells[J]. Cotton Science, 2005, 17(5): 259-263 (in Chinese).

[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] GUO Wang-zhen, Sun Jing, Zhang Tian-zhen. Gene clone and molecular breeding of cotton fiber[J]. Chinese Science Bulletin, 2003, 48(5): 411-417 (in Chinese).

[6] WILKINS T, Rajasekaran K, Anderson DM. Cotton biotechnology[J]. Crit Rev Plant Sci, 2000, 19: 511-550.

[7] ZHANG Tian-zhen. Present status and prospect of molecular breeding for cotton fiber qualities[J]. Cotton Science, 2000, 12(6): 321-326 (in Chinese).

[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]. Plant Physiol, 117: 1525.

[11] JIANG Jian-xiong, Zhang Tian-zhen. Extraction of total RNA in cotton tissues with CTAB-acidic phenolic mehtod[J]. Cotton Science, 2003, 15(3): 166-167 (in Chinese).

[12] LIU Kang, HU Feng-ping, ZHANG Tian-zhen. Effect of two methods of protein extraction from cotton ovule and fiber[J]. Cotton Science, 2005, 17(6): 323-327 (in Chinese).