Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1630 (2003) 25–34
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.)

Jinfeng Suo[a,1], Xiaoe Liang[a,1,2], Li Pu[b], Yansheng Zhang[a], Yongbiao Xue[a,*]

[a] *Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100080, China*
[b] *College of Life Sciences, Beijing Normal University, Beijing 100875, China*

Received 4 June 2003; received in revised form 21 August 2003; accepted 27 August 2003

Abstract

Cotton (*Gossypium hirsutum* L.) fibers are derived from ovule epidermis, which are developmentally similar to *Arabidopsis* trichome where several MYB transcription factors have been shown to control their formation. However, little is known about the molecular control of cotton fiber initiation. Here we isolated 55 cotton MYB domain-containing sequences expressed in ovules during fiber initiation. Among them, *GhMYB109*, encoding a R2R3 MYB transcription factor of 234 amino acids, was found to be structurally related to AtMYBGL1 and AtWER controlling the trichome initiation in *Arabidopsis thaliana*. Southern blot hybridization revealed that *GhMYB109* is present as a unique-copy gene in cotton genome. RNA expression analysis showed that it is specifically expressed in cotton fiber initial cells as well as elongating fibers. These results suggested that GhMYB109 likely plays a direct role in the initiation and elongation of cotton fiber cells.

Keywords: Cotton (*Gossypium hirsutum* L.); Fiber; MYB transcription factor; Ovule

1. Introduction

Cotton fibers are derived from single-celled, epidermal trichomes that develop from the protodermal layer of maturing seed. Fiber development usually consists of four overlapping stages, initiation, primary cell wall formation, secondary cell wall formation and maturation [1]. During the initial stages, 30% of epidermal cells (fiber initials) on the ovule surface begin to enlarge and elongate rapidly before anthesis. The primary cell wall formation starts at anthesis and lasts up to 19–20 days postanthesis (DPA). Synthesis of the secondary wall initiates about 16 DPA, overlapping with the late primary wall formation, and continues for about 40 DPA, forming a wall (5–10 μm) of almost pure cellulose [1]. Upon maturity, cotton fibers contain about 87% cellulose. Therefore, the maturation of cotton fiber takes about 40–50 days during which extensive changes in mineral content and enzyme levels/activities occur [1]. In *Arabidopsis*, several other single cell-derived structures including trichome and root hairs are developmentally similar to cotton fibers despite of their differences in shape and cell wall constitution [2,3], indicating that a similar mechanism is likely conserved for the initiation of *Arabidopsis* trichome and cotton fiber.

Compared with the *Arabidopsis* trichome, little is known about the molecular control of the cotton fiber development. So far, a number of genes differentially expressed during different stages of fiber development have been identified, but their roles in cotton development are not yet clear. Several cloned genes are fiber-specific, for example, *H6* [4], *Rac13* [5], *FbL2A* [6], *FS5* [7], *FS6* [7] and *GhEXP1* [8]. Other genes are preferentially expressed in fiber with additional expression in other tissues. For example, further expression was detected in ovary, flower and leaves for *E6* [9]; in leaves for *B6* [10] and *LTP6* [11]; and in root, flower and seed for *CelA1* [12], respectively. In addition, their expression patterns are

* Corresponding author. Tel.: +86-10-6255-2880; fax: +86-10-6253-7814.
E-mail address: ybxue@genetics.ac.cn (Y. Xue).
[1] Equal contribution.
[2] Current address: Forest Biotechnology Group, Department of Forestry, North Carolina State University, 2500 Partner II Building, Campus Box 7247, Raleigh, NC 27695, USA.

0167-4781/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbaexp.2003.08.009

usually developmentally regulated. Based on mRNA transcript accumulation, they could be divided into three groups. The first group includes those expressed in the stages of fiber elongation, such as *E6* [9], *FS6* [7], *GhEXP1* [8], *GH3* [11] and *GhCAP* [13], which peak their transcription during the elongation stage or a little earlier. The second group of the genes has the highest expression during the thickening stage of the cellulose cell wall, for example, *H6* [4] and *Fbl2A* [6], but with little expression in the early stages of fiber development. The third group has a constant expression during the complete developmental stages of fiber, for example, actins [14] and PPase [15]. Interestingly, several cotton MYB genes have been identified but none of them showed a fiber-specific expression pattern [16]. Despite of these advances, no definitive role in cotton fiber formation has been established for any of these genes.

By contrast, extensive molecular studies have revealed that a complex of transcription factors is involved in the trichome fate determination in *Arabidopsis* [23]. Among them, several MYB transcription factors play a key role in the trichome initiation [23]. *MYB* genes comprise a large family of transcription regulators in eukaryotes and are involved in a variety of biological functions. In plants, *MYB* genes are well documented [17,18] and over 100 members have been identified in *Arabidopsis* [20]. In contrast to animals, most plant *MYB* genes belong to the *R2R3-MYB* subfamily. Plant *MYB* genes have been shown to be involved in the regulation of many aspects of plant development, hormone signaling and metabolism. *AtMYBGL1* and *AtWER* are two typical *MYB* genes critical to trichome initiation in shoot and non-hair fate in root [21,22]. AtMYBGL1 is essential for trichome formation and is expressed in fields of initiating trichome cells [21], whereas AtWER is required for non-hair cell specification and is expressed in the developing non-hair cells of the root [22]. The *gl1* mutant exhibits a non-trichome phenotype, whereas *wer* mutant can induce all cell files into root hair cell files [21,22]. In addition, several other transcription factors have been shown to be required for the specification and differentiation of non-hair cells in the root and trichomes in the shoot [23]. All together, it has been proposed that the root and shoot epidermis employ a transcription factor complex that includes a MYB (WER or GL1), a bHLH (GL3 and/or an unknown bHLH) and a WD-repeat protein (TTG) to induce GL2 expression and the non-hair or trichome fate [23].

To reveal if a similar genetic control exists for *Arabidopsis* trichome and cotton fiber formation, we carried out an extensive search for MYB transcription factors expressed in cotton ovules around the fiber initiation stage in this study. As a result, a *MYB* gene similar to *AtMYBGL1* from cottonseed, *GhMYB109*, was identified. Furthermore, RNA expressional analysis showed that it is specifically expressed in the cotton fiber initials and elongating fibers.

2. Materials and methods

2.1. Plant materials

Vegetative and reproductive organs and tissues were harvested from the allotetraploid cotton species *Gossypium hirsutum* L. cv. XZ142 and *G. hirsutum* L. cv. XZ142w grown under a 30/21 °C day/night temperature regime in the greenhouse. The XZ142w is a fuzzless–lintless mutant identified by Du et al. [24] from the XZ142. Developing ovules were excised from developing flower buds or bolls on various days before/post-anthesis (DPA) relative to the day of anthesis (0 DPA).

2.2. PCR amplification of the MYB^{BRH} domain, cloning and sequences analysis

Total RNA was extracted from the mixture of −3 to 3 DPA ovule, with a Qiagen Plant RNeasy Kit (QIAGEN). Primers (Y193–201) and PCR conditions for *MYB* gene amplification were similar to those described by Romero et al. [17], and a primer list is shown in Table 1. First-strand cDNA synthesis was carried out with the Superscript II kit (Gibco-BRL). A 2-μl portion of first-strand cDNA reaction mix was used in a 50-μl PCR reaction containing 200 μM dNTPs, 1.5 mM $MgCl_2$ and 1 unit of *Taq* polymerase (Takara). Thermocycling conditions were as follows: 2 cycles of 1 min at 94 °C, 1.5 min at 41 °C, 2 min at 72 °C, 28 cycles of 1 min at 94 °C, 1.5 min at 55 °C and 2 min at 72 °C, and a final extension of 10 min at 72 °C. RT-PCR products were separated on a 1.2% agarose gel. The bands of the expected size (about 140 bp) were cloned into pGEM-T Vector (Promega) and confirmed by sequence analysis with a Beckman CEQ2000 sequencer (Beckman Coulter). cDNA fragments representing 55 *MYB* genes from cotton ovule were identified; their GenBank accession numbers are: AJ459116 to AJ459185 and *GhMYB109* is AJ549758. Alignment and

Table 1
Primers used in RT-PCR

	Name	Sequence
5′ primer	Y193	5′ cgg aat tc(a/g/t) (g/t)(a/g/t/c)a a(a/g)a g(c/t)t g(c/t)a g
	Y194	5′ cgg aat tc(a/g/t) (g/t)(a/g/t/c)a a(a/g)a g(c/t)t g(c/t)c g
	Y195	5′ cgg aat tc(a/g/t) (g/t)(a/g/t/c)a a(a/g)t c(a/g/t/c)t g(c/t)a g
	Y196	5′ cgg aat tc(a/g/t) (g/t)(a/g/t/c)a a(a/g)t c(a/g/t/c)t g(c/t)c g
	Y197	5′ ccc ggg tg(c/t) gg(a/g/t/c) aa(a/g) tc(a/g/t/c) tg
	Y198	5′ gga att ctg (c/t)gg (a/t/c/g)aa (a/g)ag (t/c)tg
3′ primer	Y199	5′ cgg aat tct t(a/g/t/c)a (c/t)(a/g/t/c)g c(a/g)t t(a/g)t c(a/g/c/t)g
	Y200	5′ cgg att tct t(a/g/t)a t(c/t)t c(a/g)t t(a/g)t c(a/t/c/g)g t
	Y201	5′ ccg aat tct t(a/g/c/t)a (c/t)(a/g/t/c)t (t/g)(a/g)t t(a/g)t c(a/g/t/c)g t

sequence identity analysis was performed with the DNAStar program. The phylogenetic tree was constructed with the ClustralW program. The Neighbor-joining method was performed based on the amino acid sequences of the MYB^{BRH} domain (BRH: the region between the conserved DNA-recognition helix [18]).

2.3. Rapid amplification of cDNA ends

The remaining sequences of *GhMYB109* were obtained by rapid amplification of cDNA ends (RACE) [19] using two cDNA-specific primers Y220 (5′-GTG GAT GAA TTA CCT GAG TCC-3′) and Y226 (5′-TCG CAA TCA AAG ACC ACC TG-3′) together with the corresponding 5′ and 3′ anchor primers (CDSIII and SMIII) (Clontech). Thermocycling conditions were as follows: 2 cycles of 1 min at 94 °C, 30 cycles of 1 min at 94 °C, 1 min at 58 °C and 2 min at 72 °C, and a final extension of 10 min at 72 °C.

2.4. Southern and Northern analyses

Genomic DNA was isolated from leaves of *G. hirsutum* L. cv. XZ142 and *G. hirsutum* L. cv. XZ142w using a cetyltrimethylammonium bromide (CTAB) extraction method [25]. The DNA (20 μg) was digested, separated on 0.8% agarose gel and transferred onto Hybond N+ (Amersham) membrane. Prehybridization, hybridization and washing of the blot were performed as recommended by the manufacturer. Total RNA was extracted from different tissues with the protocol according to RNeasy Plant Mini Kit (Qiagen). RNA samples were separated on 1.2% agarose/formaldehyde gels and transferred to Hybond N+ (Amersham) membrane, and prehybridization, hybridization and washing of the blot were performed as recommended by the manufacturer. Probes were labeled with ^{32}P by random priming using Prime-a-Gene Labeling system (Promega). The blots were exposed to X-ray films (Kodak).

2.5. In situ RNA hybridization

Digoxygenin-labeled sense or antisense RNA probes were prepared following the manufacturer's recommendation (Boehringer Mannheim). Tissue fixation and embedding, in situ hybridization and signal detection were essentially performed as described [26].

3. Results

3.1. Identification of GhMYB109 homologous to AtMYBGL1 expressed in cotton ovule

To identify R2R3 MYB genes expressed during cotton fiber development, two degenerate primers corresponding to the highly conserved MYB^{BRH} domain [17] (BRH: the region between the conserved DNA-recognition helix [18]) were used for RT-PCR amplification from cotton ovules around a fiber initiation stage (−3, 0 and 3 DPA). More than 200 clones were randomly selected for sequencing analysis. Out of them, 55 sequences were found to contain different MYB conserved domains (Fig. 1) and were named as GhMYB101 to GhMYB155. Together with the 6 *MYB* genes found by Loguerico et al. [16], the number of *MYB* gene in cotton has reached a total of 61.

To classify the cotton MYB sequences, a phylogenetic analysis was performed together with known plant MYB sequences including GhMYB1 to GhMYB6 [16] from cotton (Fig. 2). Some of them were closely related to the well-known MYB proteins in other species: GhMYB150 is closely related to two known GhMYB genes (GhMYB2 and GhMYB3 [16]) and was clustered with ZmMYBPl known to regulate anthocyanin biosynthesis [27]; GhMYB147 was clustered with PhMYB1 and AmMYBMx also known to regulate anthocyanin biosynthesis in *Petunia* [28] and cell shape of petal in *Antirrhinum* [29]; GhMYB116 was related to a virus infection induced gene-NtMYB1 [30]; while GhMYB146 was clustered with AmMYB340 and AtMYB2 known to regulate phenylpropanoid biosynthesis [31] and abscisic acid signaling [32]. Among the cotton MYB sequences isolated in this study, GhMYB109 showed the strongest similarity to the R2R3 domain of AtMYBGL1 that is a positive regulator of trichome development in *Arabidopsis* [21] (Fig. 2). To further characterize *GhMYB109*, its full-length cDNA was obtained by RACE and encodes a predicted polypeptide of 234 amino acids, representing a typical R2R3-MYB transcription factor (Fig. 3). Sequence alignment revealed that GhMYB109 has 51.2% sequence identity to AtMYBGL1 and 59.1% to AtWER. In addition, as far as the R2R3 domains were concerned, GhMYB109 showed 82.1% and 84.0% identity to AtMYBGL1 and AtWER (Table 2). Thus, GhMYB109 is more similar to AtWER than AtMYBGL1 based on their sequence comparison.

NCBI (http://www.ncbi.nlm.nih.gov) EST database search revealed that about half (28) of 55 MYB sequences identified in this study were represented by cotton ESTs. In particular, one EST of 643 bp (GA_Eb0024J14f) from a *Gossypium arboreum* 7–10 DPA fiber library was identical to *GhMYB109* cDNA, suggesting that *GhMYB109* is conserved between two cotton species. In addition, two ESTs (A1730655 and A1730139) from a 6-day *G. hirsutum* fiber cDNA library (http://www.cottongenomecentre.ucdavies.edu) were identical to *GhMYB106* and *GhMYB128*, respectively, and six were found to be highly homologous to other sequences identified in this study. Taken together, nearly half of the MYB sequences found here have not been reported previously in cotton.

3.2. GhMYB109 presents as single-copy gene in cotton genome

To examine the organization of *GhMYB109* in cotton genome, DNA blot analysis was performed using the com-

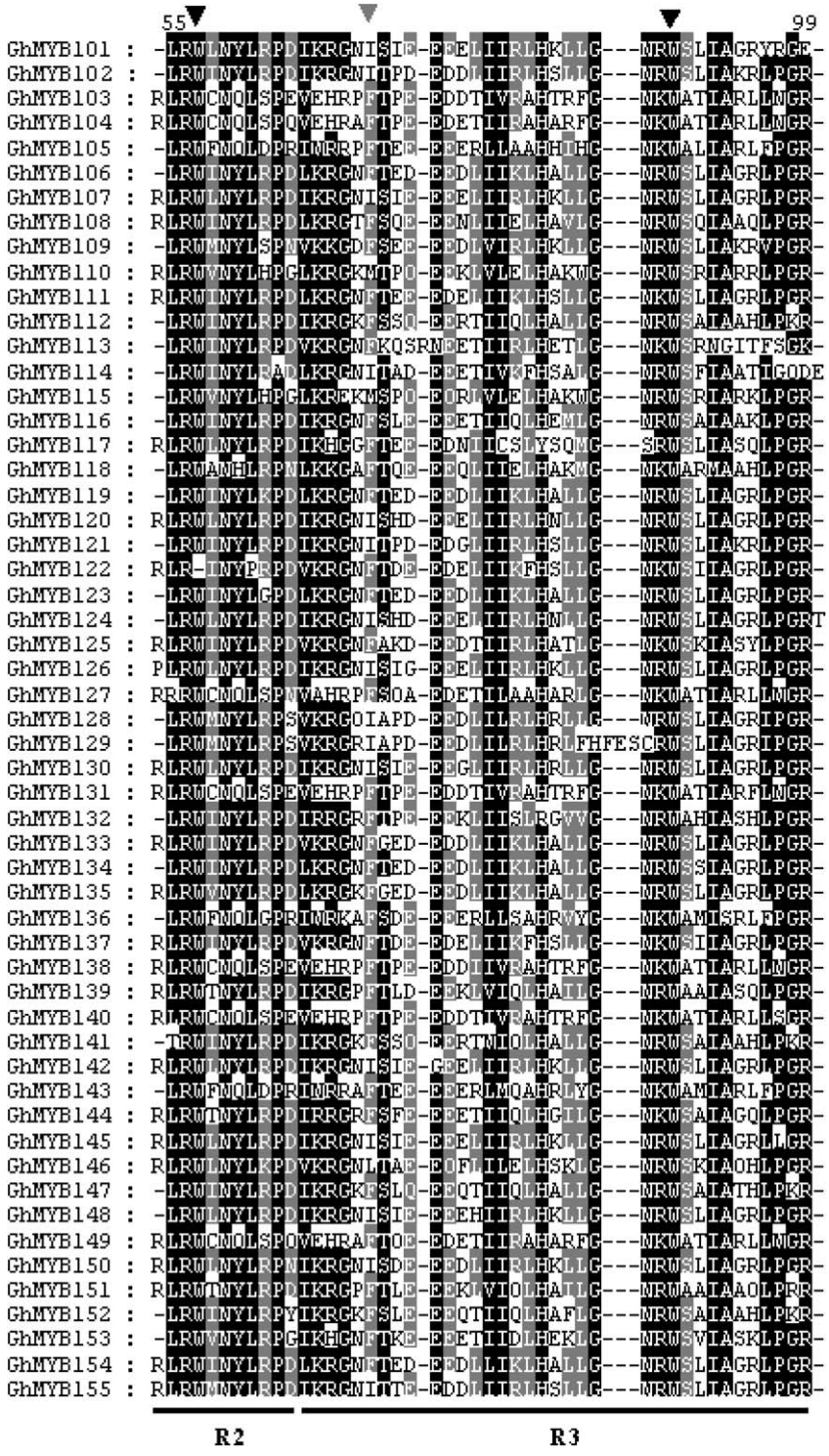

Fig. 1. Alignment of the deduced cotton MYB^{BRH} amino acid sequences. The numbers on the top of the alignment correspond to the sequence of GhMYB1 [16], and the black arrowheads indicate the conserved tryptophan residues forming a hydrophobic core. Gray arrowheads indicate aromatic amino acids that commonly substitute for the conserved tryptophan. R2 and R3 indicate the repeats in the DNA-binding domain. Identical and similar residues are shown in dark ground. Dashes indicate gaps introduced to maximize the alignment.

plete cDNA of *GhMYB109* as a probe (Fig. 4). Among the three different restriction enzymes (*Bam*HI, *Eco*RI and *Hin*dIII) used, two hybridizing bands appeared in each lane. In the *Bam*HI restriction lanes, a 5.0- and a 4.5-kb bands appeared. And the same way, a 6.5- and a 4.5-kb bands were detected in the *Eco*RI restriction lane, a 1.9- and a 1.0-kb signal showed in the *Hin*dIII's. The two hybridizing bands that appeared in each lane are consistent with *GhMYB109*'s

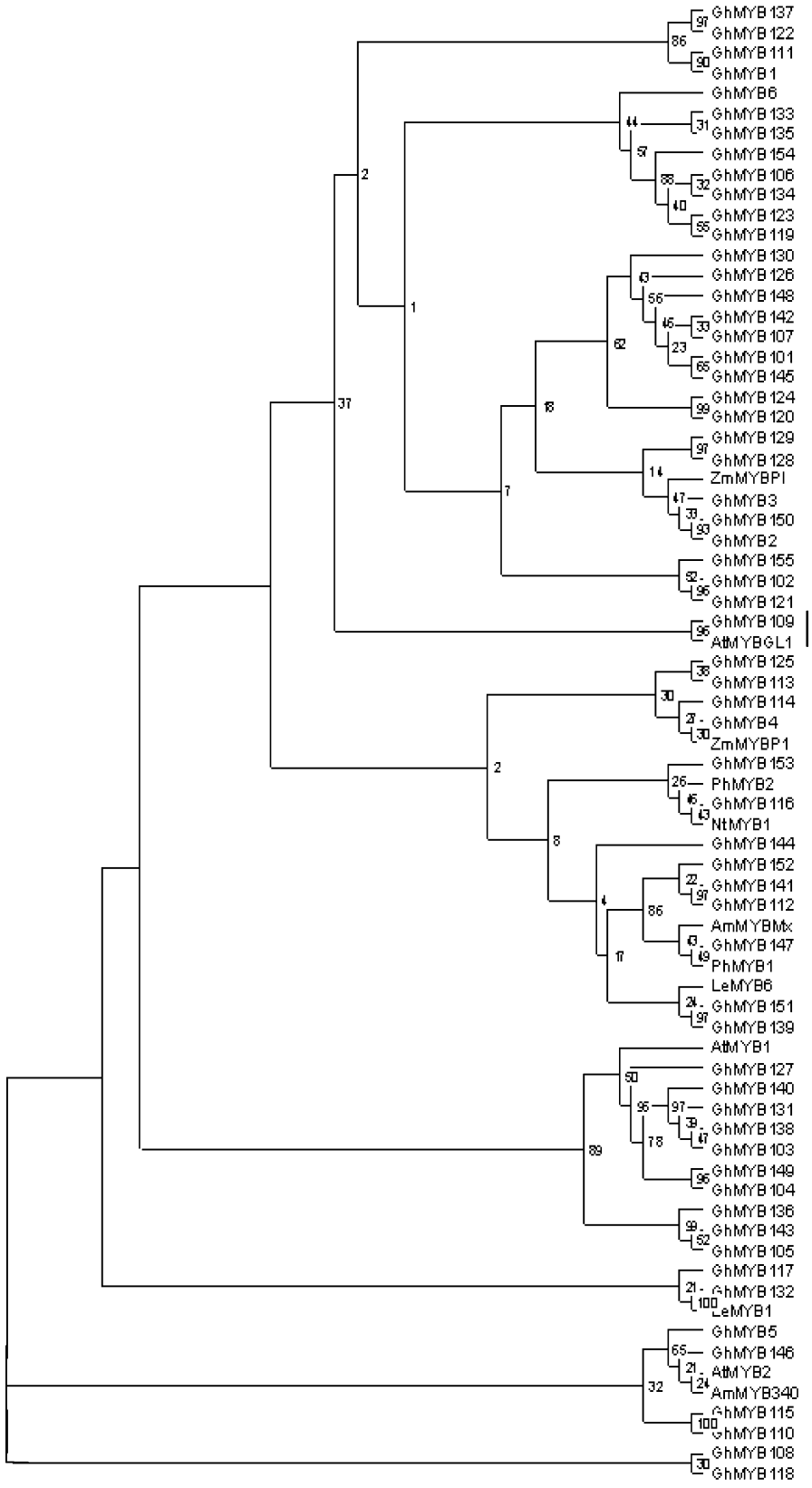

Fig. 2. A phylogenetic analysis of cotton R2R3 GhMYB sequences together with several known plant MYBs using the neighbor-joining method [41]. The tree was constructed based on the R2R3 MYB^{RBH} domains derived from the 55 partial GhMYBs isolated in this study, 6 GhMYBs [16] found previously and several plant MYB genes. They are from: *Antirrhinum majus* (AmMYB340 (JQ0959) [31]; AmMYBMx (X79108) [29]), *Arabidopsis thaliana* (AtMYBGL1 (M79448) [21]; AtMYB2 (D14712) [32]), *Lycopersicon esculentum* (LeMYB1 (X95297); LeMYB6 (X99134)), *Nicotiana tabacum* (NtMYB1 (U72762) [30]), *Petunia hybrida* (PhMYB1 (Z13996) and PhMYB2 (Z13997) [28]) and *Zea mays* (ZmMYBPl (L19494) [27]; ZmMYBP1 (Z11879)). A black line on the right side indicates GhMYB109 and AtMYBGL1.

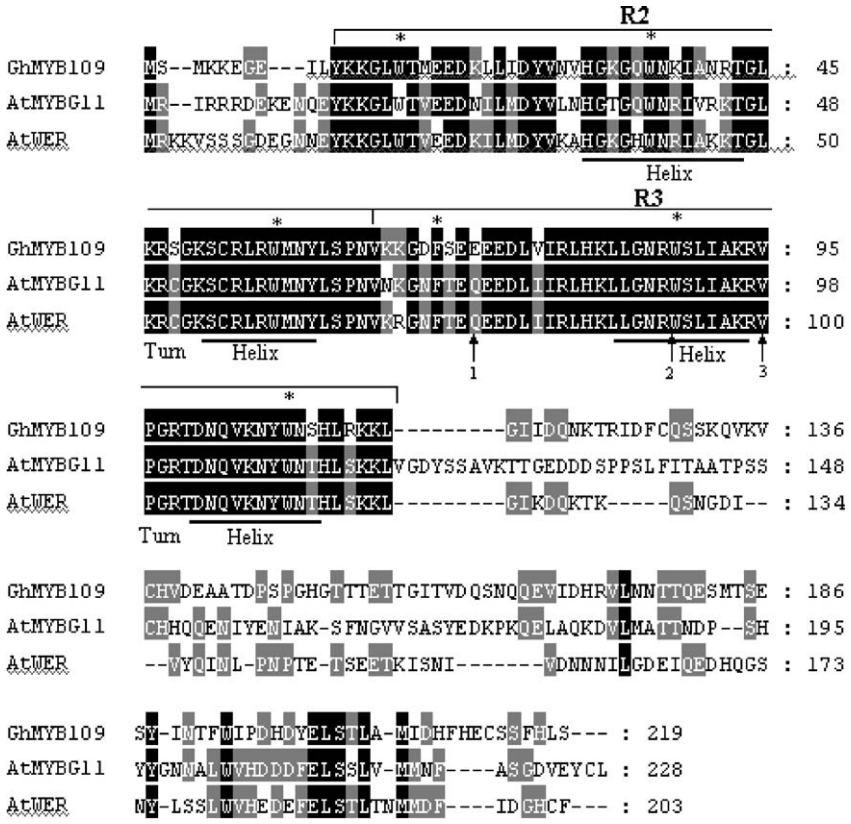

Fig. 3. Alignment of predicted polypeptide sequences of GhMYB109, AtMYBGL1 and AtWER. Identical and similar amino acids are shaded. Dashes are introduced to maximize the alignment. The helix–turn–helix motif is underlined. Regularly spaced tryptophan residues are indicated by stars. Arrows indicate the localization of introns (see Discussion).

DNA sequence which has a recognition site for all the enzymes used. To investigate if *GhMYB109* is responsible for a fiberless cotton mutant (*fl*) (see Materials and methods), DNA from the mutant was also used for DNA blot hybridization analysis. Similar bands were detected between the mutant and wild-type cotton, indicating that no gross alteration has occurred for *GhMYB109* in the fiber less mutant.

3.3. GhMYB109 is specifically expressed in cotton fiber

To examine the expression pattern of *GhMYB109*, Northern blot analysis was performed using RNA isolated from cotton fibers, ovules, anthers, petals, leaves and roots, using the complete cDNA of *GhMYB109* as a probe. The results showed that *GhMYB109* was highly expressed in fibers and moderately or weakly expressed in ovules after blossom (Fig. 5A). No expression was detected in other tissues,

Table 2
Amino acid identities among GhMYB109, AtMYBGL1 and AtWER (the numbers in parentheses indicate the identity between R2–R3 domains)

	GhMYB109	AtMYBGL1	AtWER
GhMYB109	–	51.2 (82.1)	59.1 (84.0)
AtMYBGL1	–	–	60.3 (91.5)
AtWER	–	–	–

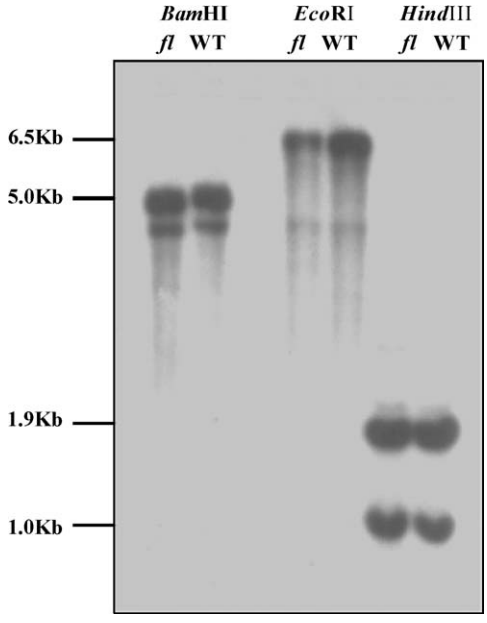

Fig. 4. DNA blot analysis of cotton genomic DNA of *GhMYB109*. Genomic DNA (20 μg/lane) of *G. hirsutum* L. cv. XZ142 (WT) and *G. hirsutum* L. cv. XZ142w (*fl*) from leaves was completely digested with *Bam*HI, *Eco*RI and *Hin*dIII, respectively, and transferred to nylon membrane. The blot was hybridized with a gene-specific probe of *GhMYB109*. Molecular weight makers are indicated in kilobases.

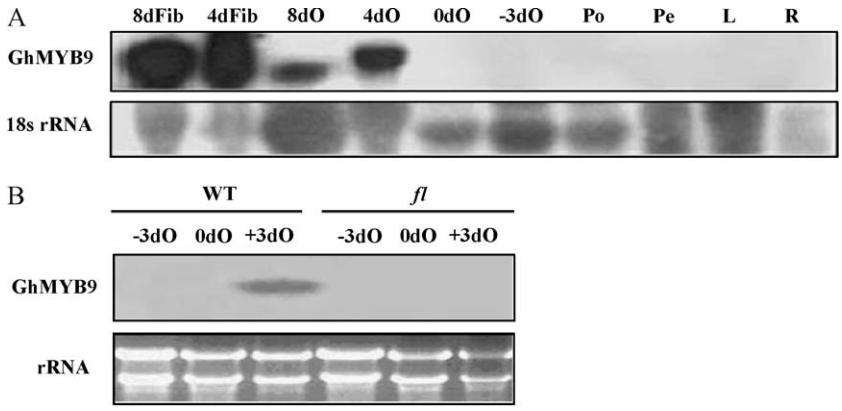

Fig. 5. RNA expression analysis of *GhMYB109*. (A) Northern blot analysis of *GhMYB109* transcripts. Total RNA (20 μg/lane) from root (R), leaf (L), petal (Pe), pollen (Po), −3, 0, 4 and 8 DPA ovules, and 4 and 8 DPA fibers was fractioned on a denaturing 1.2% (w/v) agarose gel and transferred to nylon membrane. The blot was hybridized with the full-length cDNA of *GhMYB109* and rDNA, respectively. (B) Northern analysis of *GhMYB109* transcripts between ovules from XZ142 (WT) and XZ142w (*fl*). Total RNA (10 μg/lane) from −3, 0 and 3 DPA ovule was fractioned on a denaturing 1.2% (w/v) agarose gel and hybridized with the full-length cDNA of *GhMYB109*. Loading control of total RNA was shown.

indicating that *GhMYB109* is specific to cotton ovule. To investigate whether the signal of the ovule was given by the ovule or the remaining fibers on it, the fiberless mutant with a "naked" ovule was used for Northern blot analysis. No expression was detected in the mutant ovule in contrast to the wild type (Fig. 5B), suggesting that *GhMYB109* is specifically expressed in cotton fiber. To confirm this, in situ RNA hybridization was performed using the ovules of −1, 0 and 1 DPA. An anti-sense probe of *GhMYB109* detected its transcripts in the fiber initial cells of 0 and 1 DPA ovule. But in the non-fiber cell, no signal was detected (Fig. 6). Taken together, these results clearly show that *GhMYB109* is specifically expressed in the fiber initial cells and elongating fibers.

4. Discussion

Little is known about molecular control of cotton fiber initiation and elongation. Based on the fact that both cotton

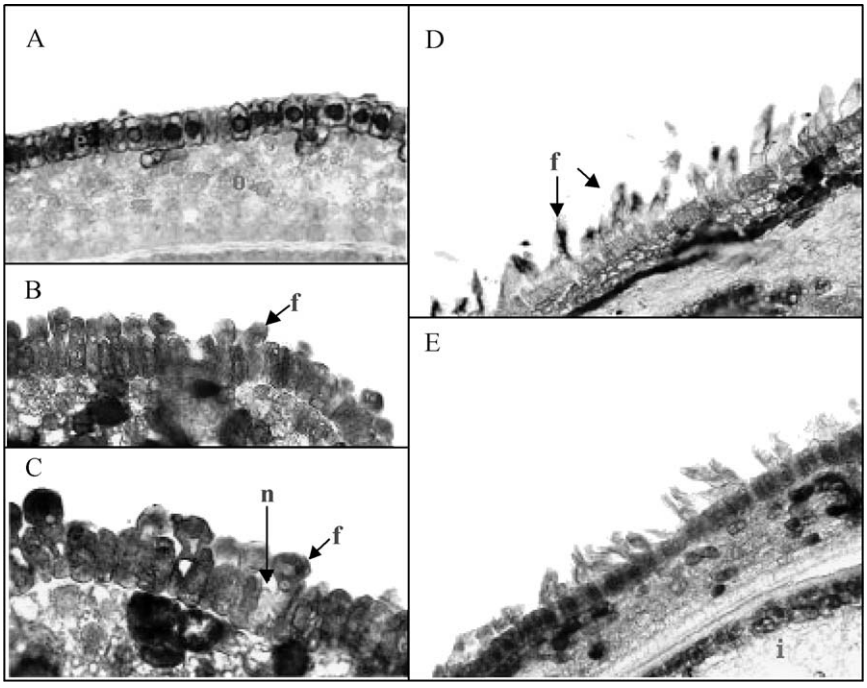

Fig. 6. In situ RNA hybridization analysis of *GhMYB109*. (A) A longitudinal section of −1 DPA ovule hybridized by an antisense probe of *GhMYB109*. (B) A longitudinal section of 0 DPA ovule hybridized by an antisense probe of *GhMYB109*. (C) An enlargement of a section in B, showing the signal only detected in the fiber initial cell but not in the non-fiber cell. (D) A longitudinal section of 1 DPA ovule hybridized by an antisense probe of *GhMYB109*, showing its expression in the elongating fibers. (E) A longitudinal section of 1 DPA ovule hybridized by a sense probe of *GhMYB109*. The blue color indicates the hybridizing signal. e: epidermal layer of cottonseed; o: outer seed coat; i: inner seed coat; f: fiber cell; n: non-fiber cell.

fiber and *Arabidopsis* trichome are derived from single epidermal cells [1,2], it is expected that a similar genetic control is likely conserved between them. In *Arabidopsis*, extensive studies have revealed that a complex of transcription factors including MYB determines the trichome fate [23]. In this study, we have isolated partial cDNA sequences corresponding to 55 different cotton *R2R3-MYB* genes expressed in the ovule during cotton fiber initiation. Among them, *GhMYB109* was shown to be specifically expressed in fiber initials and elongated fibers, suggesting that it likely plays a role in the initiation and elongation of cotton fiber.

So far, a number of cotton genes with fiber-preferential expression have been characterized, but their roles in the cotton fiber development are not yet well defined [33]. Cotton fiber formation involves the initiation of single-celled protrusions from the epidermal layer of the ovule, their subsequent linear primary growth and final secondary deposition of cellulose [1]. The majority of the genes found with a fiber-preferential expression appear to function in the stages of linear primary growth and the secondary deposition of cellulose. For example, H6 is a proline-rich protein of 21 kDa and likely belongs to the group of arabinogalactan proteins, indicating that it may be an integral part of the plasmalemma involved in the formation of the secondary wall of cotton fiber [4]. Several genes encoding α-(*GhTua 2/3*, *GhTua4*, *GhTua1* and *GhTua5*) [34] and β-tubulins (*GhTUB1*) [35] are expressed in cotton fibers and their products are thought to provide spatial information to the organization of cellulose microfibrils during cell elongation through a turgor-driven cell expansion which gives rise to a single major axis of growth. α-Expansin encoded by *GhEXP1* may play a role in cell wall extension [8]. *Ltp3* and *Ltp6* are lipid transfer protein genes associated in fiber development and possibly contribute to the cutin synthesis during the fiber primary cell wall synthesis stage [36,37]. An ACP (acyl carrier protein) cDNA predominantly expressed during the elongation stage of fiber development and might play a role in rapidly elongating cotton fibers by contributing to the synthesis of membrane lipids [13]. *GhRGP1*, encoding a putative reversibly glycosylated polypeptide (RGP), may be involved in non-cellulosic polysaccharide biosynthesis of the plant cell wall [38]. Although the promoters and other regulatory elements from E6 and *FbL2A* have been used for genetic modification of fiber properties, their identities and roles in the fiber cells are not known. Moreover, none of these genes appeared to be involved in the initiation of fiber development. Recently, a sucrose synthase gene (*Sus*) has been found to be expressed specifically in the fibers from 0 to 3 DPA seed, and its suppression led to the repression of cotton fiber cell initiation, elongation and seed development, suggesting that Sus plays a key role during these processes [39]. Nevertheless, in the transgenic seeds, the number of protrusions is still excessive, and many of them appeared to be shrunken and collapsed, suggesting that the initiation of fiber cells occurred normally but their further development was subsequently disrupted due to the lack of a sustainable *Sus* expression. Furthermore, most of these genes can be loosely described as fiber specific, because many ‘‘fiber-specific’’ genes, like *GhTUB1*, also exhibit low levels of expression in other tissues. In addition, none of the above genes expressed specifically in cotton fibers appears to encode a transcription factor. By contrast, GhMYB109 is specifically expressed in the cotton fiber initials and elongating fibers, indicating that it might play a direct role in the initiation and elongation stages of the fiber cells.

GhMYB109 is closely related to AtMYBGL1 and AtWER (Fig. 3, Table 2), further supporting its postulated role in cotton fiber formation. Although several cotton *MYB* genes have been isolated previously, none of them showed an expression pattern specific to fiber [16]. *AtMYBGL1* and *AtWER* genes encode related proteins of the R2R3 MYB class of transcriptional regulators and have been demonstrated to control the initiation of the single cell-based trichome and root hair development in *Arabidospsis*, respectively [23]. Despite their distinct developmental functions, AtWER and AtMYBGL1 proteins are functionally equivalent [23]. However, their expression profiles are quite different. AtWER is a position-dependent regulator and is expressed only in the non-hair cells of root; whereas the outcome in the shoot is unpredictable and likely influenced by stochastic fluctuations in the concentration of transcriptional regulators [23]. Thus, the expression pattern of *GhMYB109* appears more like *AtMYBGL1* than that of *AtWER*, indicating that molecular control is likely conserved between trichome and cotton fiber initiation. In *Arabidopsis*, MYB proteins can be divided into three groups by the different intron/exon structure [17]. Group A (accounting for about 10% of the *A. thaliana* proteins), which also includes the animal and protist R2R3-MYB proteins, represents genes with no intron in the RBH region. Group B (5% of the *A. thaliana* proteins) represents proteins encoded by genes with an intron at position 3 (Fig. 3). Finally, group C (85% of *A. thaliana* proteins) contains genes with an intron at position 2 (Fig. 3) [17]. AtMYBGL1 belongs to group C, whereas GhMYB109 is a protein of group A, which has no intron in the BRH region. However, both GhMYB109 and GhMYBGL1 have an intron in the same position in R2 region. It is generally believed that most introns are gained during a long period of evolution and the intron insertion is an ongoing process [40]. The presence of an additional intron in AtMYBGL1 indicates that it likely evolved later than GhMYB109. Furthermore, *AtMYBGL1* is expressed only in the initiation stage of trichome [21], but the transcripts of GhMYB109 are detected from the cotton initials through to 8-day-old fibers. These results also support the role of GhMYB109 both in the initiation and elongation of cotton fiber formation, and the latter function may have had lost for AtMYBGL1 in trichome development.

However, it is not known if GhMYB109 is a true orthologue of AtMYBGL1 because we could not exclude

the possibility that cotton fiber expresses another factor(s) whose function(s) is more similar to AtMYBGL1. Currently, we are testing if GhMYB109 is functionally similar to AtMYBGL1 or AtWER by complementing their mutants in *Arabidopsis*. In addition, we are also searching for other cotton factors interacting with GhMYB109 through a yeast two-hybrid screening.

Acknowledgements

The project is supported by grants from the Ministry of Science and Technology of China (2001AA222051) and the Chinese Academy of Sciences. We thank Dr. Xiongming Du from Cotton Research Institute of Chinese Academy of Agricultural Sciences for providing cotton *fl* mutants.

References

[1] A.S. Basra, C.P. Malik, Development of the cotton fiber, Int. Rev. Cytol. 89 (1984) 65–113.
[2] M. Hülskamp, S. Miséra, G. Jürgens, Genetic dissection of trichome cell development in *Arabidopsis*, Cell 76 (1994) 555–566.
[3] J. Van't Hof, Increased nuclear DNA content in developing cotton fiber cells, Am. J. Bot. 86 (1999) 776–779.
[4] M.E. John, G. Keller, Characterization of mRNA for a proline-rich protein of cotton fiber, Plant Physiol. 108 (1995) 669–676.
[5] D.P. Delmer, J.R. Pear, A. Andrawis, D.M. Stalker, Genes encoding small GTP-binding proteins analogous mammalian rac are preferentially expressed in developing cotton fibers, Mol. Gen. Genet. 248 (1995) 43–51.
[6] J.A. Rinehart, M.W. Petersen, M.E. John, Tissue-specific and developmental regulation of cotton gene FbL2A. Demonstration of promoter activity in transgenic plants, Plant Physiol. 112 (1996) 1331–1341.
[7] S.J. Orford, J.N. Timmis, Abundant mRNAs specific to the developing cotton fibre, Theor. Appl. Genet. 94 (1997) 909–918.
[8] S.J. Orford, J.N. Timmis, Specific expression of an expansin gene during elongation of cotton fibers, Biochim. Biophys. Acta 1398 (1998) 342–346.
[9] M.E. John, L.J. Crow, Gene expression in cotton (*Gossipium hirsutum* L.) fiber: cloning of the mRNAs, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 5769–5773.
[10] M.E. John, Characterization of a cotton (*Gossipium hirsutum* L.) fiber mRNA (Fb-B6), Plant Physiol. 107 (1995) 1477–1478.
[11] D.P. Ma, H.C. Liu, H. Tan, R.G. Creech, J.N. Jenkins, Y.F. Chang, Cloning and characterization of a cotton lipid transfer protein gene specifically expressed in fiber cells, Biochim. Biophys. Acta 1344 (1997) 111–114.
[12] J.R. Pear, Y. Kawagoe, W.E. Schreckengost, D.P. Delmer, D.M. Stalker, Higher plants contain homologs of the bacterial CelA genes encoding the catalytic subunit of cellulose synthase, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 12637–12642.
[13] P. Song, R.D. Allen, Identification of a cotton fiber-specific acyl carrier protein cDNA by differential display, Biochim. Biophys. Acta 1351 (1997) 305–312.
[14] Y. Shimizu, S. Aotsuka, O. Hasegawa, T. Kawada, T. Sakuno, F. Sakai, T. Hayashi, Changes in levels of mRNAs for cell wall-related enzymes in growing cotton fiber cells, Plant Cell Physiol. 38 (1997) 375–378.
[15] L.B. Smart, F. Vojdani, M. Maeshima, T.A. Wilkins, Genes involved in osmoregulation during turgor-driven cell expansion of developing cotton fibers are differentially regulated, Plant Physiol. 116 (1998) 1539–1549.
[16] L.L. Loguerico, J.Q. Zhang, T.A. Wilkins, Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.), Mol. Gen. Genet. 261 (1999) 660–671.
[17] I. Romero, A. Fuertes, M.J. Benito, J.M. Malpica, A. Leyva, J. Paz-Ares, More than 80R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*, Plant J. 14 (1998) 273–284.
[18] P.D. Rabinowicz, E.L. Braun, A.D. Wolfe, B. Bowen, E. Grotewold, Maize R2R3 *Myb* genes: sequence analysis reveals amplification in the higher plants, Genetics 153 (1999) 427–444.
[19] M.A. Frohman, M.K. Dush, G.R. Martin, Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer, Proc. Natl. Acad. Sci. U. S. A. 85 (1988) 8998–9002.
[20] The *Arabidopsis* Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, Nature 408 (2000) 796–815.
[21] J.C. Larkin, D.G. Oppenheimer, S. Pollock, M.D. Marks, *Arabidopsis GLABROUS1* gene requires downstream sequences for function, Plant Cell 5 (1993) 1739–1748.
[22] M.M. Lee, J. Schiefelbein, WEREWOLF, a MYB-related protein in *Arabidopsis*, is a position-dependent regulator of epidermal cell patterning, Cell 99 (1999) 473–483.
[23] J. Schiefelbein, Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot, Curr. Opin. Plant Biol. 6 (2003) 74–78.
[24] X.M. Du, J.J. Pan, R.H. Wang, T.Z. Zhang, Y.Z. Shi, Genetic analysis of presence and absence of lint and fuzz in cotton, Plant Breed. 120 (2001) 519–522.
[25] A.H. Paterson, C.L. Brubaker, J.F. Wendel, A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis, Plant Mol. Biol. Report 11 (1993) 122–127.
[26] Z. Lai, W. Ma, B. Han, L. Liang, Y. Zhang, Y. Xue, An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum, Plant Mol. Biol. 50 (2002) 29–42.
[27] K.C. Cone, S.M. Cocciolone, F.A. Burr, B. Burr, Maize anthocyanin regulatory gene *pl* is a duplicate of *c1* that functions in the plant, Plant Cell 5 (1993) 1795–1805.
[28] J. Avila, C. Nieto, L. Canas, M.J. Benito, J. Paz-Ares, *Petunia hybrida* genes related to the maize regulatory C1 gene and to animal myb proto-oncogenes, Plant J. 3 (1993) 553–562.
[29] K. Noda, B.J. Glover, P. Linstead, C. Martin, Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor, Nature 369 (1994) 661–664.
[30] Y. Yang, D.F. Klessig, Isolation and characterization of a tobacco mosaic virus-inducible myb oncogene homolog from tobacco, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 14972–14977.
[31] D. Jackson, F. Culianez-Macia, A.G. Prescott, K. Roberts, C. Martin, Expression patterns of myb genes from *Antirrhinum* flowers, Plant Cell 3 (1991) 115–125.
[32] H. Abe, T. Urao, T. Ito, M. Seki, K. Shinozaki, K. Yamaguchi-Shinozaki, *Arabidopsis* AtMYC2 (bHLH) and AtMYB2 (MYB) function as transcriptional activators in abscisic acid signaling, Plant Cell 15 (2003) 63–78.
[33] S. Ji, Y. Lu, J. Feng, G. Wei, J. Li, Y. Shi, Q. Fu, D. Liu, J. Luo, Y. Zhu, Isolation and analyses of genes preferentially expressed during early cotton fiber development by subtractive PCR and cDNA array, Nucleic Acids Res. 31 (2003) 2534–2543.
[34] D.J. Whittaker, B.A. Triplett, Gene-specific changes in alpha-tubulin transcript accumulation in developing cotton fibers, Plant Physiol. 121 (1999) 181–188.
[35] X. Li, L. Cai, N. Cheng, J. Liu, Molecular characterization of the cotton *GhTUB1* gene that is preferentially expressed in fiber, Plant Physiol. 130 (2002) 666–674.

[36] C.Y. Hsu, R.G. Creech, J.N. Jenkins, D.P. Ma, Analysis of promoter activity of cotton lipid transfer protein gene *LTP6* in transgenic tobacco plants, Plant Sci. 143 (1999) 63–70.
[37] H.C. Liu, R.G. Creech, J.N. Jenkins, D.P. Ma, Cloning and promoter analysis of the cotton lipid transfer protein gene Ltp3, Biochim. Biophys. Acta 1487 (2000) 106–111.
[38] G. Zhao, J. Liu, Isolation of a cotton RGP gene: a homolog of reversibly glycosylated polypeptide highly expressed during fiber development, Biochim. Biophys. Acta 1574 (2002) 370–374.
[39] Y.L. Ruan, D.J. Llewellyn, R.T. Furbank, Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development, Plant Cell 15 (2003) 952–964.
[40] L. Duret, Why do genes have introns? Recombination might add a new piece to the puzzle, Trends Genet. 17 (2001) 172–175.
[41] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.