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

Jinfeng Suoa,1, Xiaoe Lianga,1,2, Li Pu, Yansheng Zhanga, Yongbiao Xuea,*

aInstitute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100080, China
bCollege 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.

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
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 FhI2A [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 Arabidopsis 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 cell specification 55 representing 55 members of the MYB family of plant development, hormone signaling and metabolism.

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, GmMYB109, was identified. Furthermore, RNA expression 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<sup>BRH</sup> 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<sub>2</sub> 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
sequence identity analysis was performed with the DNASTar program. The phylogenetic tree was constructed with the ClustalW program. The Neighbor-joining method was performed based on the amino acid sequences of the MYB\textsuperscript{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 cetyl-trimethylammonium 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 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

Digoxigenin-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\textsuperscript{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 ZmMYBPI 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 AnMYB340 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-
plete cDNA of GhMYB109 as a probe (Fig. 4). Among the three different restriction enzymes (BamHI, EcoRI and HindIII) used, two hybridizing bands appeared in each lane. In the BamHI 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 EcoRI restriction lane, a 1.9- and a 1.0-kb signal showed in the HindIII’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<sup>RBH</sup> 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 (Q0959) [31]; AmMYBMx (X79108) [29]), Arabidopsis thaliana (AtMYBGL1 (M79448) [21]; AtMYB2 (D14712) [32]), Lycopersicon esculentum (LeMYB1 (X95297); LeMYB6 (X99134)), Nicotiana tabacum (NmMYB1 (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.
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 fiberless 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

<table>
<thead>
<tr>
<th></th>
<th>GhMYB109</th>
<th>AtMYBGL1</th>
<th>AtWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>GhMYB109</td>
<td>–</td>
<td>51.2 (82.1)</td>
<td>59.1 (84.0)</td>
</tr>
<tr>
<td>AtMYBGL1</td>
<td>–</td>
<td>–</td>
<td>60.3 (91.5)</td>
</tr>
<tr>
<td>AtWER</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

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).

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 BamHI, EcoRI, and HindIII, respectively, and transferred to nylon membrane. The blot was hybridized with a gene-specific probe of GhMYB109. Molecular weight makers are indicated in kilobases.
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. 5. RNA expression analysis of GhMYB109.](image)

![Fig. 6. In situ RNA hybridization analysis of GhMYB109.](image)
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 *Arabidopsis*, 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 RBH 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. Xiaoming Du from Cotton Research Institute of Chinese Academy of Agricultural Sciences for providing cotton fl mutants.

References


