Plant Molecular Biology **51:** 313–325, 2003.

Evolution and expression of *MYB* genes in diploid and polyploid cotton

M. L. Cedroni[1†], R.C. Cronn[2†], K.L. Adams[†], T.A. Wilkins[‡] and J.F. Wendel[*†]
[†]*Department of Botany, Iowa State University, Ames, IA 50011* [‡]*Department of Agronomy & Range Science, University of California-Davis, Davis, CA 95616 Department of Botany, Iowa State University, Ames, IA 50011; Fax: 515-294-1337,* [1]*Current Address: Genome Sequencing Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park, Saint Louis, MO 63108* [2]*Current Address: U.S. Forest Service, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331; (*author for correspondence: e-mail jfw@iastate.edu*

Received 4 February 2002; accepted in final form 4 June 2002

Key words: cotton; duplicate gene expression; evolution; gene family; *MYB*; polyploidy

Abstract

R2R3-MYB transcription factors have been implicated in a diversity of plant-specific processes. Among the functions attributed to myb factors is the determination of cell shape, including regulation of trichome length and density. Because myb transcription factors are likely to play a role in cotton fiber development, the molecular evolutionary properties of six *MYB* genes previously shown to be expressed in cotton fiber initiation were examined. In accordance with their presumed central role, each of the genes display conservative substitution patterns and limited sequence divergence in diploid members of the genus *Gossypium*, and this pattern is conserved in allotetraploid cottons. In contrast to highly reiterated rDNA repeats, *GhMYB* homologues (duplicated gene pairs) exhibit no evidence of concerted evolution, but instead appear to evolve independently in the allopolyploid nucleus. Expression patterns for the *MYB* genes were examined in several organs to determine if there have been changes in expression patterns between the diploids (*G. raimondii* and *G. arboreum*) and the tetraploid (*G. hirsutum*) or between the duplicated copies in the tetraploid. Spatial and temporal expression patterns appear to have been evolutionarily conserved, both during divergence of the diploid parents of allopolyploid cotton and following polyploid formation. However, the duplicated copies of *MYB1* in the tetraploid are not expressed at equal levels or equivalently in all organs, suggesting possible functional differentiation.

Introduction

The genus *Gossypium* includes the world's most important fiber crop plant, namely, cotton. There are about 50 members in the genus, ranging from shrubs to small trees in arid regions throughout the tropics and subtropics (Wendel, 1995). Most of the genus is diploid (n = 13), but five species from the western hemisphere are classic genomic allopolyploids ('AD-genome'; n = 26), formed from hybridization between two diploid species ('A-genome' and 'D-genome'), possibly in the mid-Pleistocene (Wendel, 1989; Seelanan *et al.*, 1997; Cronn *et al.*, 2002). Four different species have been independently domesticated at different times in human history, including the New World allopolyploids *G. hirsutum* L. and *G. barbadense* L. and the Old World diploids *G. herbaceum* L., and*G. arboreum* L. (Brubaker *et al.*, 1999). Currently, over 90% of the world's cotton derives from *G. hirsutum*, while *G. barbadense* produces most of the remaining commercial cotton.

Cotton 'fiber' used in textiles derives from single-celled, epidermal trichomes that develop from the protodermal layer of maturing seed. Seed trichome features vary widely in the genus, ranging from short, colored, and tightly adherent in most wild species to long, fine and white in the cultivated species (Fryxell, 1979; Applequist *et al.*, 2001). The genes involved in the determination of fiber initiation, elongation, secondary cell wall synthesis and maturity are largely

unknown, although a fiber model has been generated based upon available expression profiles (Wilkins and Jernstedt, 1999). Although relatively few studies have directly demonstrated a connection between genetic change and the evolution of novel plant phenotypes, many authors have suggested that transcription factors that regulate plant-specific processes may play a significant role in morphological evolution (Baum,1998; Cubas *et al.*, 2001; Doebley and Lukens, 1998; Luo *et al.*, 1999; Purugganan, 1998).

Recent work has demonstrated that several members of the *MYB* family of transcription factors are expressed in cotton ovules during fiber initiation and expansion (Loguercio *et al.*, 1999). Using the highly conserved DNA-binding domain (DBD), six members of the *R2R3-MYB* family of transcription factors were isolated from a developing cotton ovule cDNA library. Expression analysis determined that the six *MYB* genes are differentially regulated during ovule and/or fiber development. This initial study has prompted a more thorough effort to understand the role of *MYB* genes in morphological evolution of the cotton fiber, and the importance of *R2R3*-MYB factors in plant-specific processes. The *R2R3-MYB* family of transcription factors is one of the largest regulatory gene families known in plants (Martin and Paz-Ares, 1997; Romero *et al.*, 1998; Kranz *et al.*, 1998; Jin and Martin, 1999; Rabinowicz *et al.*, 1999; Reichmann *et al.*, 2000). While myb factors have been found in all organisms studied to date, the number of *MYB* genes found in plants is significantly higher than in other organisms. The origin and evolutionary history of members of this large gene family are thus of interest.

MYB factors are defined by the helix-turn-helix motif of their DNA-binding domain. In animals, most MYB proteins contain three imperfect repeats of this domain (R1, R2, and R3), while in plants most MYB proteins have only two repeats (R2 and R3), although both one repeat and three repeat mybs have been found (Kranz *et al.*, 2000; Braun and Grotewold, 1999). Evidence from maize and *Arabidopsis* places the number of *MYB* genes at over 125 in both genomes (Romero *et al.*, 1998; Meissner *et al.*, 1999; Rabinowicz *et al.*, 1999; Stracke *et al.*, 2001). Recent unpublished surveys of in sorghum (Jiang and Peterson, unpubl.) and cotton (Cedroni and Wendel, unpubl.) suggest that there may be closer to 200 *MYB* genes in each of these crop species. The *R2R3-MYB* family of transcription factors have been implicated in a diversity of functions, including influencing cell identity and cell fate (Payne *et al.*,1999; Wada *et al.*, 1997), polysaccharide secretion during seed germination (Penfield *et al.*, 2001), inflorescence development (Kirik *et al.*, 1998), mediation of gibberellic acid signaling during flowering (Gocal *et al.*, 2001), and regulation of tryptophan (Bender and Fink, 1998) and flavonoid pigment synthesis (Noda *et al.*, 1994; Borevitz *et al.*, 2000; Zhang *et al.*, 2000). This evidence indicates that *R2R3-MYB* genes play a significant role in developmental processes in higher plants (Jin and Martin, 1999; Stracke *et al.*, 2001)

To understand the genetic basis of cotton fiber development we have been exploring the diversity, evolution and expression of *R2R3-MYB* genes. In this report, we explore the molecular evolutionary properties of the six previously described (Loguercio *et al.*, 1999) cotton *MYB* genes, using the well-known phylogenetic history of the cotton genus as a framework for comparison. Of particular interest was whether there was evidence of directional selection for any of the six *MYB* genes and whether rates of sequence divergence have been equivalent among the diploid and polyploid genomes (cf. Cronn *et al.*, 1999). Expression patterns were also explored to evaluate the effects of polyploidy on duplicate gene expression and possible tissue-specific partitioning of function, as well as to assess whether expression patterns were conserved during divergence at the diploid level. These latter questions were motivated by the possibility of divergence in function following gene duplication, as suggested by theory (Force *et al.*, 1999; Lynch and Force, 2000; Wendel, 2000).

Materials and methods

Plant materials

Six cotton *GhMYB* genes (*GhMYB1* through *GhMYB6*; Loguercio *et al.*, 1999) were isolated from 4 different species, the allopolyploid *G. hirsutum* L. (AD-genome; accession TM1), living models of its two ancestral genome donors, *G. herbaceum* L. (A-genome; accession A1-73) and *G. raimondii* Ulbrich (D-genome; 'Galau's'), and *Gossypoides kirkii* (Masters) J.B. Hutchinson, which was included as a phylogenetic outgroup (Seelanan *et al.*, 1997). For one gene (*GhMYB4*) it was necessary to isolate one of the two duplicated copies (from the A genome) from a BAC clone from a *Gossypium barbadense* Pima library (AD-genome). For expression analysis, organs

were harvested from three different taxa, *G. raimondii*, *G. arboreum* L. cv. AKA8401 (A-genome), and *G. hirsutum* L cv. TM1. Young leaves were harvested from greenhouse-grown plants and processed on the same day, cotyledons were harvested from 7-day old seedlings, and all floral organs (petals, androecium, and gynoecium) were harvested on the day of anthesis. Developing cotton ovules and attached fibers were harvested 5, 10, 15, and 20 days post-anthesis (dpa).

Gene amplification and sequencing

Based on the work of Loguercio *et al.* (1999) and sequences deposited in GenBank, forward primers were designed for PCR-amplification of full-length (or nearly full-length) sequences from each of the six cotton fiber *MYB* genes (*GhMYB1-6*; Figure 1). Forward primers used were as follows: Myb1F: GTAACTGATGGGACGATCACCTTGTT; Myb2F: GGCTAATAATGGCTCCAAAGAAGGCTG; Myb3F: GGGCCACTAAAGAATGGAGCA; Myb4F: GGCAGTTACACATAGGGAGAT; Myb5F: GCCTCTCCGACTGTAATTAACC; Myb6F: GAAACTCCGATGAGAAAACCTTGCT. With the exception of *GhMYB4*, reverse primers described in Loguercio *et al.* (1999) were used. To isolate *GhMYB4*, a reverse primer was designed to amplify the entire coding region (Myb4R: TTTGTAAACTGATTTGTGATGGA). Standard PCR reactions were done in 25 μl volumes containing 1 unit *Taq* polymerase (Gibco), 1× PCR buffer (Gibco), 2 mM $MgCl_2$, 200 μM of each dNTP, and 40 μM of primer in an MJ Research PTC-100 thermal cycler (Watertown, Massachussets). PCR cycling parameters were 1 min at 80 °C followed by 30 cycles of 1 min at 95 °C, 1 min at 50 °C, 2 min at 72 °C and then a final 10 min extension at 72 °C. PCR products from the diploid species for *GhMYB1, -2, -3*, and *-5* were directly sequenced following gel isolation with QIAquick kit (Qiagen Inc., Valencia, CA). To confirm sequences and check problematic regions arising from direct sequencing of PCR amplicons, *GhMYB4* and *GhMYB6* isolated from *G. herbaceum* and *G. raimondii* were cloned into pGemT (Promega Corp., Madison, WI) according to the manufacturer's instructions, and sequenced from plasmid DNA templates.

Because the allopolyploid contains two copies ('homologues') of each gene derived from A-genome and D-genome diploid progenitors, it was necessary to clone the amplification pools for each *MYB* gene studied in *G. hirsutum*. Using the same forward/reverse primer-pairs, PCR products were also cloned into pGemT. Homologues were identified by diagnostic restriction site analysis prior to sequencing, or by comparison to sequences from diploids. To avoid possible complications due to PCR recombination in allopolyploid cotton (Cronn *et al.*, 2002), for each *GhMYB* gene a minimum of three clones was used to determine the sequence for each duplicated copy. In cases where PCR recombination was evident, 10 or more clones were sequenced to ensure that the correct sequence was obtained (Cronn *et al.*, 2002). Due to primer bias for the D-genome of *G. hirsutum* in the cloning of *GhMYB4*, the A-genome homologue was obtained by sequencing a BAC clone (159-H20) from a *G. barbadense* Pima library. BAC DNA was isolated using 50ml cultures and the Psi Clone Big BAC DNA Extraction Kit (Princeton Separation, Inc; Adelphia, NJ).

Southern hybridization analysis

Southern blot analysis was used to verify that each of the six *GhMYB* homologs studied was single-copy in the taxa surveyed. RFLP and hybridization methods used to identify gene copy number and mapping polymorphisms in diploid and tetraploid cottons followed previously described methods (Brubaker *et al.*, 1999; Reinisch *et al.*, 1994). Briefly, five micrograms of genomic DNA from two A-genome species (*G. herbaceum* A_1-73; *G. arboreum* A_2-47), two D-genome species (*G. raimondii*, *G. trilobum*), and two AD-genome species (*G. hirsutum* race Palmeri, *G. barbadense* K101) was evaluated for RFLP variation using at least five restriction enzymes, including *Eco*RI, *Hin*dIII, *Eco*RV, *Bam*HI, and *Pst*I (*Dra*I and *Cfo*I were also included in A-genome surveys). Gene-specific hybridization probes were generated by PCR amplification using gene-specific primers (Fig. 1). Probes were labeled by PCR amplification, using the Ambion Strip-EZ kit with ^{32}P-dATP and hybridized using stringent hybridization conditions (final wash in 0.5× SSC, 0.1% SDS at 65 °C).

Phylogenetic analyses and determination of substitution rates

For each of the six *R2R3-MYB* genes isolated from five genomes (A- and D-genome *Gossypium*, *Gossypiodes kirkii*, and the A and D homologues from the allotetraploid), phylogenetic analysis was conducted using PAUP* 4.0b1 (Swofford, 1998). Because of the small number of genes in each analysis (N = 5), maximum

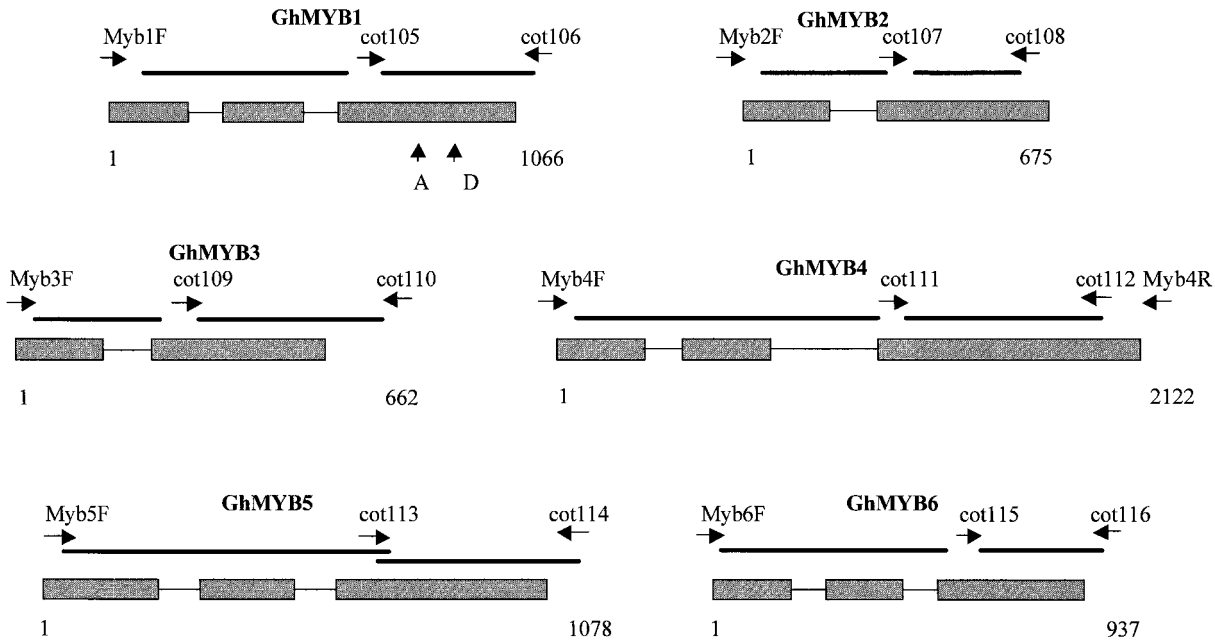


Figure 1. Diagram of the six *GhMYB* genes studied. Exons and introns are shown as boxes and lines, respectively. The highly conserved R2R3 regions and the regions amplified for use as hybridization probes are represented by thicker lines above each gene (left and right lines, respectively). Arrows above indicate names and locations of primers. The probe regions are highly divergent from one another and hence provide gene specificity under stringent wash conditions. For each gene the probe region was also amplified for analysis of transcript accumulation patterns, except for *GhMYB2* for which the larger amplification product was used. Arrows under *GhMyb1* denote location of genome-diagnostic restriction sites. Aligned length, including UTRs, is provided beneath each figure.

parsimony analysis was performed using exhaustive searches. For each of the trees recovered, inferred substitutions were inspected for evidence of intergenomic recombination in the allopolyploid nucleus and topologies were inspected for branch length inequalities. The former would suggest gene conversion or some other form of molecular interaction among homologues, whereas the latter would suggest relaxed or strong directional selection. Rates of synonymous and nonsynonymous substitutions were calculated for the coding region of each *MYB* using DnaSP version 3 (Rozas and Rozas, 1999).

RNA Isolation and RT-PCR

Total RNA was isolated from *G. arboreum, G. raimondii*, and *G. hirsutum* using published methods (Wilkins and Smart, 1996) modified for small-scale extractions. Total RNA was extracted from 0.2–0.3 g (fresh weight) from the androecium, the gynoecium (not including ovules), cotelydons, leaves, and petals. Additionally, RNA was extracted from ovules collected from single bolls at 5, 10, 15, and 20 days after anthesis. To remove trace contaminants of genomic DNA, RNA was treated with DNase I according to manufacturer's instructions (Ambion, Austin, TX). Following DNAse treatment, RNA concentration was determined spectrophotometrically (A260). First-stand cDNA syntheses for all organs except ovules included approximately 2 μg of total RNA and was accomplished using the two-step Ambion RETROscript Kit. To evaluate expression in ovules, equal amounts of ovular RNA from 5, 10, 15, and 20 dpa (0.5 μg each) were added to first strand syntheses to yield a cDNA pool that covered early stages of cotton fiber initiation (e.g., 5 dpa), to the later expansion phase (e.g, 10–20 dpa). PCR cycling conditions and gene-specific primers (Fig. 1) detailed in the previous section were used to amplify *GhMYB* transcripts for each of the six genes from each organ type and taxon. Genomic DNA was also used in separate reactions to provide positive controls.

Results from RT-PCR were used to determine if patterns of *MYB* transcript accumulation were conserved between diploid and tetraploid species. To determine if the A and D homologues of *G. hirsutum* are both expressed, *GhMYB1* PCR products were digested with diagnostic restriction enzymes that discriminate between homologous copies in *G. hirsutum*. The al-

lotetraploid A-genome homologue was diagnosed by a unique restriction site for *Nla*III, while the D-genome copy was distinguished by a unique *Pflm*I site.

Single-stranded conformation polymorphism (SSCP) analysis

RT-PCR products of *GhMYB1* were gel purified using the QiaQuick kit (Qiagen) and labeled with ^{33}P as follows: 4 ng of RT-PCR product, 0.5 μM reverse primer, 2.5 mM $MgCl_2$, 0.2 mM of each cold dNTP, 0.8 mCi ^{33}P-dCTP, and *Taq* polymerase were mixed in 10 μl reaction volumes and cycled in a MJ PTC-100 thermal cycler for 2 min at 94 °C followed by 30 cycles of 30 s at 94 °C, 30 cycles at 57 °C, 1 min at 72 °C and then a final 6 min extension at 72 °C. Labeled RT-PCR products were electrophoresed through mutation enhancement detection gels (FMC) containing 10% weight/volume urea (Yip *et al.*, 1999) and 0.6 × Tris-Taurine-EDTA ('TTE', Amersham) running buffer. Band quantitation was accomplished using a Molecular Dynamics Storm 840 phorphorimager and ImageQuant software.

Results

Characterization of six cotton R2R3-MYB genes

Homologues for six *R2R3-MYB* genes were PCR-amplified from the genomes of three diploid species, *G. raimondii, G. herbaceum*, and *Gossypoides kirkii* (the latter representing the outgroup taxon). Additionally, homoeologous copies representing the A and D genomes of tetraploid cottons were obtained and sequenced from *G. hirsutum*, except that *GhMYB4* for the A-genome of allopolyploid cotton was obtained using a BAC from *G. barbadense* (see *Materials and methods*, above). Southern hybridization of genomic blots indicate that the*R2R3-MYB* genes studied are each represented by a single locus in diploids and two homologous loci in tetraploids (Loguercio *et al.*, 1999 and data not shown). These results lend confidence to the supposition that orthologous copies of the *GhMYB* genes were isolated from the diploids, as well as both homologous copies from allotetraploid cotton.

Gossypium R2R3-MYB genes are typical of the majority of *R2R3-MYB* genes described, having the same intron/exon structure (Kranz *et al.*, 1998; Romero *et al.*, 1998) and consensus amino acid sequence for the DNA binding domain common to plant *R2R3-MYBs* (Martin and Paz-Ares, 1997). Full-length gene sequences for five of the six *GhMYB* genes were obtained, while for *GhMYB2*, the sequence was complete except for 87 bp at the 3′ end of the gene (Table 1; Figure 1). The six genes studied ranged in length from 662 bp (*GhMYB3*) to approximately 2160 bp (*GhMYB4*). The length of *GhMYB6* was not precisely determined due to strings of AT-rich sequences (12 to 15 nucleotides in length) within both introns that precluded determination of a clear consensus sequence, even after multiple sequencing attempts. No variation in exon length was observed within *Gossypium* for any of the genes, but exon III of *GhMYB1* and *GhMYB5* in *Gossypioides kirkii* were two and one codons shorter, respectively, than their *Gossypium* orthologues. Accordingly, length variation among orthologues for any of the genes was due to insertion/deletion events within introns, primarily intron II (Figure 1). This intron, present in most angiosperm *MYB* genes, typically ranges from 75 to 100 bp in length. However, intron II is absent in orthologues of *GhMYB2* and *GhMYB3*, and is considerably longer (629 bp) in orthologues of *GhMYB4*. Although the absence of intron II from *GhMYB2* and *GhMYB3* makes these genes somewhat unusual, approximately 10% of all angiosperm *R2R3-MYB* genes lack this intron (Romero *et al.*, 1998). In *GhMYB4*, both introns are significantly longer than the corresponding regions in the other five cotton *MYB* genes. Intron II ranges from 82 bp in *GhMYB5* to 629 bp in *GhMYB4*, and exon III ranges from 421 bp in *GhMYB6* to 1107 bp in *GhMYB4*, as reported in Loguercio *et al.* (1999).

At the protein level, the six cotton *R2R3-MYB* factors contained motifs highly conserved among plant *R2R3-MYB*s (Kranz *et al.*, 1998; Martin *et al.*, 1997; Romero *et al.*, 1998). Based on inter-genic comparisons of amino acid and nucleotide substitutions, *GhMYB5* is the most divergent of the cotton *MYB*s. Analysis of amino acid sequences reveals that even in the most highly conserved regions, *GhMYB5* differs more from the other five genes than any other intergenic comparisons (data not shown).

Phylogenetic analysis of the six cotton R2R3-GhMYB *genes*

Phylogenetic analysis of the *GhMYB* genes revealed topologies expected from the well-established organismal history (e.g., Cronn *et al.*, 1999). Specifically, when rooted with the orthologous sequence from the outgroup *Gossypioides kirkii*, *Gossypium* sequences for each of the genes fall into two clades, each contain-

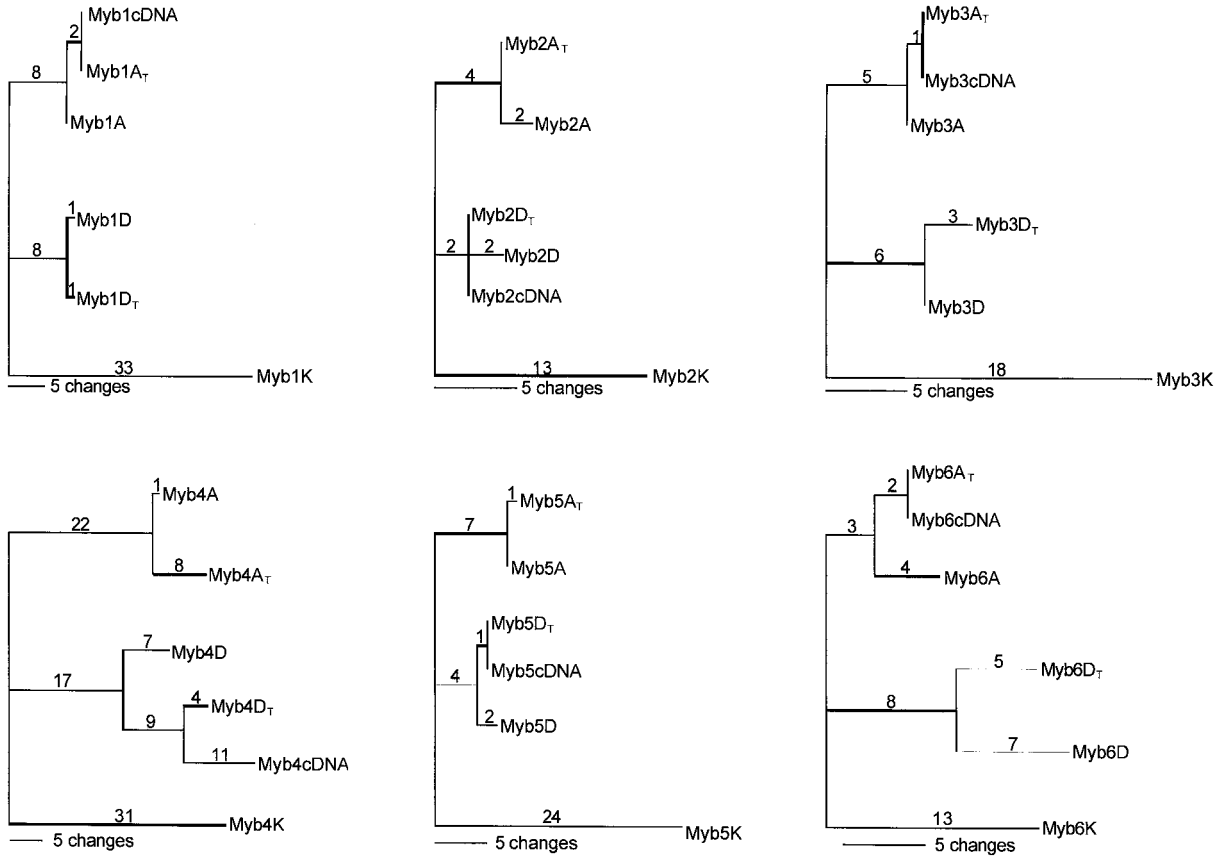


Figure 2. Phylogenetic analysis of six *GhMYB* genes expressed during fiber development in cotton. Included for each gene are genomic sequences from diploid (*G. herbaceum*, 'A'; *G. raimondii*, 'D') and allopolyploid (A_T and D_T) *Gossypium*, as well as the cDNA clones from *G. hirsutum* reported in Loguercio *et al.* (1999). *Gossypiodes kirkii* ('k') was included as a phylogenetic outgroup. Branch lengths indicate the number of inferred nucleotide changes.

ing one of the two homoeologues from allopolyploid cotton and the corresponding copy from the progenitor diploid genome (Figure 2). This topology was obtained in all six cases, demonstrating that duplicated genes in the allopolyploid have evolved independently of one another following polyploid formation, as reported previously for other homoeologous locus-pairs in *Gossypium* (Cronn *et al.*, 1999). There was little to no phylogenetic character conflict within each data set (data sets for *GhMYB1* through *GhMYB5* had zero homoplasy, and retention and consistency indices were 0.93 and 0.79, respectively, for *GhMYB6*), and thus we judge each topology to be robust.

For each of the six genes studied, one or more cDNA sequences from *G. hirsutum* were available for comparison (Loguercio *et al.*, 1999). Except for *GhMYB4*, the cDNA sequences were identical to one of the homologues recovered from *G. hirsutum* genomic DNA. Inspection of the *GhMYB4* gene tree indicates that while the cDNA is similar to that obtained from the D genome of the allopolyploid, it is more divergent than expected, suggesting it is a closely related but distinct gene (paralogue). This suggestion is supported by other evidence as well. Specifically, difficulties encountered in sequencing *GhMYB4* from *G. raimondii* and *G. herbaceum* necessitated cloning of PCR products; sequencing of five clones from *G. herbaceum* (A-genome) yielded at least two different sequences. While Southern blots suggested that *GhMYB4* exists as a single-copy gene per diploid genome, sequence analysis of multiple clones resulting from PCR amplification in both A-genome and D-genome diploids reveals a degree of variability greater than expected from heterozygosity at a single locus (Figure 2). Hence, it may be that *GhMYB4* underwent a gene duplication event prior to divergence of A- and D-genome diploid cottons.

Table 1. Characterization of six *GhMYB* genes in diploid and allopolyploid cotton.

Locus	Aligned length	Absolute length	Exon lengths	Intron 1 length	Intron 2 length	GenBank numbers
GhMYB1	1066	1060 (A_T)	E1 = 133	86	89 (A_T	AY115501
		1063 (D_r	E2 = 130	89 (k)	92 (D_r	AY115502
		1063 (D_5)	E3 = 622		92(D_5)	AY115503
		1060 (A_1)	616 (k)		89 (A_1)	AY115504
		1059(k)			91 (k)	AY115505
GhMYB2[a]	588	675 (A_T	E1 = 136	78 (A_T	absent	AY115506
	(675)	674 (D_T	E2 = 461	78 (D_T		AY115507
		674 (D_5)		77 (D_5)		AY115508
		675 (A_1)		77 (A_1)		AY115509
		674 (k)		77 (k)		AY115510
GhMYB3	662	662	E1 = 133	80 (A_T)	absent	AF37708
			E2 = 449	80 (D_T		AF377307
				80 (D_5)		AF377316
				80 (A_1)		AF377318
		661 (k)		79 (k)		AY115511
GhMYB 4[b]	2122	? (A_T	E1 = 133	102	? (A_T)	n.a.
		2068 (D_T	E2 = 130		596 (D_T)	AY115522
		2066 (D_5)	E3 = 949		594 (D_5)	AY115523
		2092 (A_1)	1099 (cDNA)		629 (A_1)	AY115524
		? (k)			? (k)	n.a.
GhMYB 5[d]	1078	1064 (A_T)	E1 = 160	100	82 (A_T	AY115512
		1071 (D_T	E2 = 130	95 (k)	89 (D_T)	AY115513
		1070 (D_5)	E3 = 592		88 (D_5)	AY115513
		1064 (A_1)	589 (k)		82 (A_1)	AY115515
		1070 (k)			96 (k)	AY115516
GhMYB 6[c]	933-937	921-925 (A_T	E1 = 130	92-25 (A_T	152-153 (A_T)	AY115517
		930-934 (D_T	E3 = 130	95-98 (D_T	154-155 (D_T	AY115518
		925-929 (D_5)	E3 = 421	91-94 (D_5)	154-155 (D_5)	AY115519
		924-928 (A_1)		92-95 (A_1)	152-153 (A_1)	AY115520
		893-897 (k)		87-90 (k)	129-130 (k)	AY115521

Taxa include an allotetraploid, the two genome donors to the allotetraploid, and an outgroup. A_T = A genome from the alloptetraploid *G. hirsutum*; D_T = D genome from *G. hirsutum*; D_5 = *G. raimondii*; A_1 = *G. herbaceum*; k = *Gossypiodes kirkii*. Lengths indicated are conserved unless parenthetically noted otherwise with taxon abbreviations.

[a]Partial sequence used in this analysis

[b]Portions of the A genome and *Gossypiodes kirkii* genes were not obtained

[c]The range of values for *GhMYB6* reflect strings of Ts that were not unambiguously resolved.

[d]Aligned length (cDNA) differs from Loguercio *et al.* (1999) due to an extra base in the original sequence.

Rates of sequence evolution

Nucleotide substitution rates were calculated for all pairwise comparisons among orthologous copies for each *GhMYB* gene and were tabulated separately for synonymous (*Ks*) and nonsynonymous (*Ka*) sites (Table 2). Due to problems encountered in sequencing *GhMYB4*, only partial sequences were obtained from *Gossypiodes kirkii* and the A-genome homologue (A_T) from *G. hirsutum*. The region of the gene amplified for these two species includes most of intron II and part of exon III. As these are the most variable regions within *MYB* genes, *GhMYB4* was excluded from the following calculations.

Table 2. Synonymous and nonsynonymous substitution rates for *GhMYB* genes in cotton[a]

	Ks/Ka/Ks:Ka ratio[c]					
Locus	D_5 vs A_1	D_T vs A_T	A_1 vs A_T	D_5 vs D_T	Comparisons to k[d]	Total Replacements
GhMYB1	0.0429	0.0490	0.0059	0.0000	0.1030	E1 = 1
	0.0082	0.0099	0.0016	0.0033	0.128	E2 = 0
	5:2:1	4.9:1	3.7:1	–	8.1:1	E3 = 11
GhMYB2[b]	0.0426	0.0381	0.0104	0.0000	0.0343	E1 = 1
	0.0081	0.0027	0.0000	0.0054	0.0259	E2 = 10
	5.3:1	14.1:1	–	–	1.3:1	
GhMYB3	0.0278	0.0469	0.0091	0.0091	0.0764	E1 = 1
	0.0094	0.0141	0.0000	0.0047	0.0201	E2 = 11
	3.0:1	3.3:1	–	2.0:1	3.8:1	
GhMYB5	0.0270	0.0215	0.0000	0.0053	0.0451	E1 = 4
	0.0090	0.0105	0.0015	0.0000	0.0185	E2 = 2
	3.0:1	2.0:1	–	–	2.4:1	E3 = 12
GhMYB6	0.0144	0.0217	0.0071	0.0000	0.0163	E1 = 0
	0.0186	0.0144	0.0082	0.0123	0.0176	E2 = 0
	0.8:1	1.5:1	0.9:1	–	0.9:1	E3 = 17

[a]Taxa include an allotetraploid, the two genome donors to the allotetraploid, and an outgroup. A_T = A genome from the allotetraploid *G. hirsutum*; D_T = D genome from *G. hirsutum*; D_5 = *G. raimondii*; A_1 = *G. herbaceum*; k = *Gossypiodes kirkii*. *GhMYB4* was omitted because of incomplete data from the allopolyploid A genome and *Gossypiodes kirkii*. E2 (exonII) in *GhMYB2* and *GhMYB3* is equivalent to exonII + exonIII in other *GhMYB* genes.

[b]Partial sequence used for this analysis

[c]For each comparison, there are three numbers that represent (top) the number of synonymous substitutions per synonymous site (Ks); (middle) the number of nonsynonymous substitutions per nonsynonymous site (Ka); (bottom) the Ks:Ka ratio

[d]Mean of all pairwise comparisons between A_1, D_5, A_T, D_T

The nucleotide substitution data reinforce the phylogenetic results in showing that rates of sequence substitution are modest. As shown in Table 2, synonymous rates vary approximately three-fold among the *GhMYB* genes, with the slowest evolving gene being *GhMYB6* and the fastest being *GhMYB1* and *GhMYB2*. When mean divergence between all ingroup sequences and the outgroup orthologue are calculated for each gene, this range increases to nearly six-fold, due to the particularly long branch separating the *GhMYB1* sequences from the *Gossypioides kirkii* sequence. The data also underscore the high degree of evolutionary constraint placed on the *R2R3-MYB* genes, as relatively few replacement substitutions were detected, and most Ks/Ka ratios are relatively high (Table 2). The noteworthy exception is for *GhMYB6*, where the Ks/Ka ratio approaches unity. Inspection of substitution patterns indicated there were only four silent mutations in the coding region compared to 17 nonsynonymous substitutions, all in exon III and distributed without any apparent pattern among the five sequences studied.

Examination of branch lengths shown in Figure 2 reveals a high degree of similarity between sequences from the diploid genomes and their counterparts (orthologues) in the allopolyploid, as expected from a relatively recent origin of the polyploids (Wendel, 1989). Branch lengths were otherwise unremarkable, suggesting approximately equal rates of nucleotide substitution at the diploid and allopolyploid levels, as has been shown for most other cotton genes (Cronn *et al.*, 1999; Small and Wendel, 2000; see, however, Small *et al.*, 1998). Formal tests of evolutionary rate equivalence conducted using the Tajima 1D and 2D tests (Tajima, 1993) confirm this observation.

Rates of sequence evolution were compared at the diploid and allopolyploid levels. If rates of sequence evolution are independent of ploidy, then the amount of divergence between orthologues from diploids (D vs. A) should be the same as that between homologues

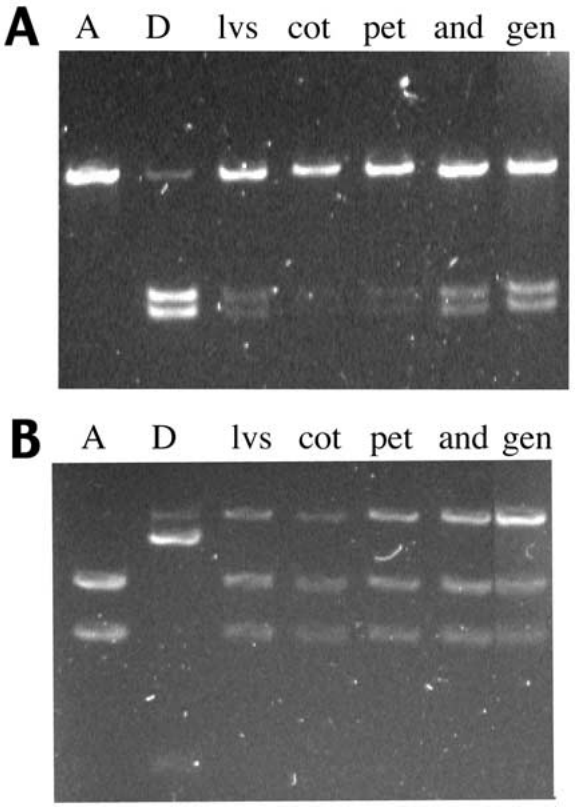


Figure 3. Expression patterns of homologous *GhMYB1* copies assayed by restriction digestion. The first two lanes are for the diploids ('A' = *G. arboreum*; 'D' = *G. raimondii*) while the remaining lanes are from allopolyploid *G. hirsutum* ('lvs' = leaves; 'cot' = cotyledons; 'pet' = petals; 'and' = androecium; and 'gen' = genomic DNA as a control). **Panel A**. Restriction digest using *Pfl*M1. This enzyme has a unique restriction site in *GhMYB1* in the D diploid and in the corresponding genome of allopolyploid cotton. **Panel B.** Restriction digest using *Nla*III. This enzyme reveals a diagnostic, shared restriction site within the A diploid and the corresponding band of allopolyploid cotton. There is no *Nla*III site in the D genome copy from allopolyploid cotton, but there is a site unique to the D diploid.

(D_T vs. A_T) of tetraploids. As shown in Table 2, this appears to be the case; in two of five comparisons (*GhMYB2* and *GhMYB5*), K_s values of diploid sequences exceed that of homologues from the allopolyploid, whereas for the other three genes, the reverse is true. The fluctuation in diploid/polyploid comparisons likely reflects stochastic variation associated with relatively low divergences, and we conclude that there is no evidence of accelerated sequence evolution for this sample of *R2R3-MYB* genes from allopolyploid cotton.

Conservation of expression in diploids and the tetraploid

Patterns of transcript accumulation for the six *GhMYB* genes for tetraploid cotton were explored by Loguercio *et al.* (1999). This analysis is extended here to show that transcript accumulation patterns for the six *GhMYB* genes generally appear to be evolutionarily conserved between diploids and allopolyploids. For each locus, RT-PCR results revealed that all three species (*G. arboreum*, *G. raimondii*, and *G. hirsutum*) shared expression patterns for all organs examined (results not shown). *GhMYB1*, *GhMYB2*, and *GhMYB6* are expressed in leaves, cotyledons, petals, and the androecium, whereas *GhMYB3*, *GhMYB4*, and *GhMYB5* show a more restricted distribution of transcript accumulation. *GhMYB3* and *GhMYB4* transcripts were detected only in the gynoecium, while *GhMYB5* was not detected in any of the organs tested (*contra* Loguercio *et al.* 1999).

Relative expression of GhMYB1 *genes duplicated by polyploidy*

To evaluate whether the *GhMYB1* mRNAs detected by RT-PCR experiments reflect transcription of one or both homologues in the allopolyploid, we used two methods to evaluate the relative contribution of each tetraploid genome to the overall transcript pool in leaves, cotyledons, petals, and the androecium. The first method used restriction enzymes diagnostic of each of the two homologues to selectively digest each homoeolog (Song and Osborn 1994). *GhMYB1* was the only gene that showed a relatively high level of expression in all five organs tested (the four mention above, plus gynoecium), and for which appropriately useful diagnostic restriction sites existed. A 538 bp region of *MYB1* (representing base pairs 531–1069) was examined for restriction enzyme sites diagnostic of each of the two homologues and their corresponding diploid antecedants (Fig. 3). The allotetraploid A-genome homoeologue was diagnosed by a unique restriction site for *Nla*III that is shared by the A diploid but not the D-genome homologue, while the D-genome copy was distinguished by a unique *Pflm*I site that is shared by the D diploid but not the A-genome homologue. *GhMYB1* RT-PCR products were digested with *Nla*III or *Pflm*I and the fragments were separated by agarose gel electrophoresis. Genomic PCR products of *GhMYB1* were digested for use as a control; the copies from both genomes are present

in equal amounts in genomic DNA (Figure 3). RT-PCR products from the A diploid and the D diploid were used as size controls. Digestion of RT-PCR and PCR products from the A-genome homologue or the A diploid with *Pflm*I yielded a product of 538 bp, and digestion of the D-genome homologue or the D diploid yielded products of 275 bp and 263 bp (Figure 3A). Digestion of RT-PCR and PCR products from the A-genome homologue or the A diploid with *Nla* III yielded products of 220 bp and 318 bp, and digestion of the D-genome homologue or the D diploid yielded a product of 538 bp (Figure 3B). As judged by qualitative examination of the RT-PCR gels (Figure 3), the *GhMYB1* transcript pool does not appear to be equally derived from the two homologues; instead, there is a biased accumulation of A-genome transcripts, particularly in leaves, cotyledons, and petals.

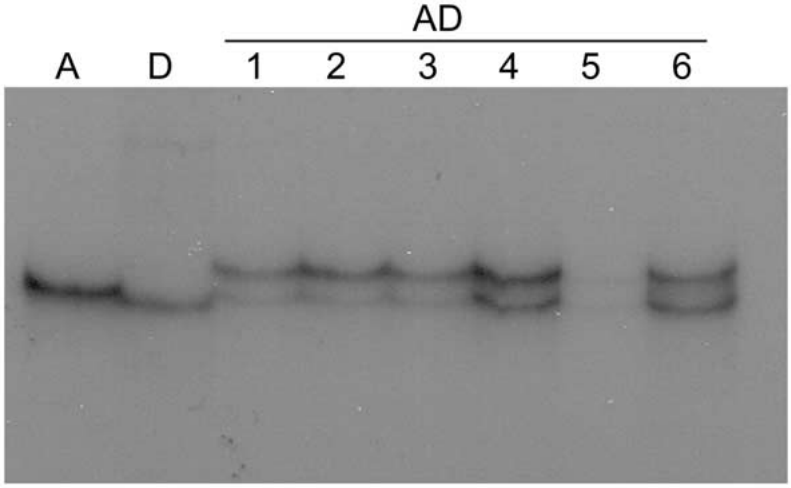
To further evaluate the relative contribution of each tetraploid genome to the overall *GhMYB1* transcript pool, single-stranded conformation polymorphism (SSCP) analysis (Orita *et al.*, 1989) was performed to separate the two homologous *GhMYB1* gene transcripts and determine their proportions. SSCP analysis has been used to detect mutations in members of gene families by exploiting electrophoretic mobility differences due to altered conformations (e.g., Slabaugh *et al.*, 1997). RT-PCR products were separated by conformation using mutation detection enhancement gels, as described in the *Materials and methods*. Figure 4 shows that *GhMYB1* transcripts from the A genome are more abundant than transcripts from the D genome in leaves (69:31), cotyledons (75:25), petals (68:32), and the androecium (63:37); transcript levels from the gynoecium (lane 5) are not reported due to low signal. The results of the restriction digestions and SSCP analysis indicate that *GhMYB1* homologues are differentially and spatially regulated and that the most abundant genome-specific transcripts depend on the organ-type.

Discussion

Cronn *et al.* (1999) demonstrated that most genes duplicated during polyploidy in *Gossypium* are expected to exhibit sequence patterns indicative of independent evolution in the allopolyploid nucleus. The molecular evolution of the cotton *R2R3-MYB* genes studied confirmed these expectations. Five of the *MYB* genes (all but *MYB4*) exist as single-copy genes in each progenitor diploid genome and in a pattern indicative of functional conservation. For each of the *GhMYB* genes, phylogenetic analysis yields the topology expected from the known history of *Gossypium*, in the process providing additional evidence that the homoeologous copies are evolving independently subsequent to allopolyploid formation. Moreover, there appears to be a general equivalence in evolutionary rates at the diploid and allopolyploid levels. Though these results confirm conclusions reached earlier (Cronn *et al.*, 1999), they provide a contrast to the non-independent evolution exhibited by homoeologous ribosomal genes (Wendel *et al.*, 1995), which are highly reiterated in the genome. Thus, the present study offers a new twist in that while each of the *MYB* genes is part of a large *R2R3-MYB* gene family, their evolutionary behavior resembles that of single-copy genes.

Patterns of sequence substitution for *GhMYB6* warrant further comment. The number of nonsynonymous substitutions ($n = 17$) is much greater than the number of silent mutations in the coding region ($n = 4$). These results might be expected in the case of strong directional selection, but this is not evident in the phylogenetic analysis (Figure 2) or substitution data (Table 2), neither of which demonstrate rate acceleration that might be indicative of selection. Additionally, it does not appear as though the actual number of replacements is unusually high for *MYB6*, since the number of replacements ranges from 12 to 16 for the other five genes. Therefore, it seems plausible to suggest that the aberrant Ks:Ka ratio reflects a decrease in the rate of silent substitution for *MYB6*. The recent study of Hurst and Pal (2001) is notable in this respect, in suggesting that selection may act to decrease the number of silent mutations, perhaps through strong codon usage bias. Whether this type of selection applies to *MYB6* is presently unknown.

In addition to conservation of copy number and molecular evolutionary rates for the genes, expression patterns also appear mostly to be conserved among diploids and between diploid and allopolyploid *Gossypium*. Furthermore, at least in the case of *GhMYB1*, detection of genome-specific gene transcripts clearly indicates that both homologues are expressed. However, in some organ types (leaves, cotyledons, petals, and androecium) the A genome transcripts are more abundant than the D genome transcripts, suggesting the possibility of nascent partitioning of function of duplicated genes in the polyploid. Recent work (Wilkins *et al.*, unpubl.) has established that *GhMYB1* functions as a novel repressor in sec-



Tissue	A genome	D genome
1. Leaf	69.1 ± 2.2	32.4 ± 2.2
2. Cotyledon	75.8 ± 1.4	24.2 ± 1.4
3. Petal	67.6 ± 2.6	32.4 ± 2.6
4. Androecium	62.8 ± 2.0	37.2 ± 2.0
6. genomic DNA	51.9 ± 2.6	48.1 ± 2.6

Figure 4. Expression patterns of homologous *GhMYB1* copies assayed by single-stranded conformational polymorphism analysis. A = *G. arboreum*: D = *G. raimondii*: and AD = *G. hirsutum*. Although transcript migration in SSCP gels from the diploid species does not correspond to that of their counterparts in the allopolyploid, fragment isolation followed by amplification and sequence analysis demonstrates that the upper and lower bands in allopolyploid cotton (lanes 1-6) correspond to alternative homologues. Percentages of each transcript pool are shown below (with phosphorimager standard deviations based on 10 replicates), along with organ types.

ondary metabolism with affects on phenylpropanoid and lignin biosynthesis. Perhaps there has been functional divergence in repressor activity following allopolyploid formation and the attendant gene duplication. It will be interesting to expand the analysis to include other duplicated *GhMYB* genes, to explore the possibilities of partitioning of ancestral function, or 'subfunctionalization' (Force *et al.*, 1999; Lynch and Force, 2000), as well as proportional contributions of the two homologues to the transcriptome.

In their initial work, Loguercio *et al.* (1999) discovered that each of the six *GhMYB* genes is developmentally and spatially regulated in *G. hirsutum*, implying that each of these genes encodes a functional protein. There is no evidence, either at the nucleotide or amino acid levels, to suggest that any of the genes has become nonfunctional in any of the taxa studied here. In fact, analysis of the pattern of nucleotide substitutions and the RT-PCR evidence to date suggests that each of these genes is expressed, in at least one organ, both in the diploids *G. raimondii* (D-genome) and *G. arboreum* (A-genome), and in the derived allopolyploid *G. hirsutum* (AD genome).

Since the R2R3 region of *MYB* genes represents the DNA binding domain, the high conservation in this region (Martin and Paz-Ares, 1997; Kranz *et al.*, 1998; Romero *et al.*, 1998) is readily explained by functional constraints. It seems probable, however, that the remaining portions of *MYB* genes are not under as stringent stabilizing selection. This is reflected not only in the size variation in exon III (or exon II in those genes lacking intron II), but also in the patterns of sequence substitution, where exon III apparently is able to tolerate a higher frequency of nonsynonymous substitutions (Table 2). The C-terminal portion of R2R3*MYB* genes contains the trans-activation domain and is known to be highly variable, with only a few conserved motifs identified. The trans-activation domain mediates protein-protein interactions, so variability here may influence the types of protein interactions and varying affinities for interacting partners, and hence its regulation of target gene expression.

GhMYB4 presents a likely exception to our initial hypothesis that each of the six cotton *MYB* genes is present in only one copy per diploid genome. As noted previously, two unique sequences were found for both the A-genome diploid species *G. herbaceum* and the D-genome of the allotetraploid *G. hirsutum*. Southern hybridization results do not indicate that *GhMYB4* exists in more than one copy per diploid genome, but these data do not exclude the possibility of a recent duplication where flanking restriction sites are conserved. Alternatively, *GhMYB4* may be heterozygous in both species in question, although these species reproduce primarily by self-pollination and heterozygosity is rarely evident (Wendel *et al.*, 1989; Brubaker and Wendel, 1994). The detection of a probable recent gene duplication for an *R2R3-MYB* transcription factor echoes the historical massive expansion of the R2R3 MYB gene family in plants and indicates that the process remains active.

Acknowledgements

We thank Andy Paterson and Rick Noyes for providing BAC clones used for isolation of *GhMYB* genes from allopolyploid cotton, and Tammy Haselkorn and Corrinne Grover for technical assistance. Financial support was provided by the National Science Foundation, the United States Department of Agriculture, the US-Israel Binational Science Foundation, The Jane Coffin Childs Memorial Fund for Medical Research, and the Plant Science Institute of Iowa State University.

References

Applequist, W.L., Cronn, R.C., Wendel, J.F., 2001. Comparative development of fiber in wild and cultivated cotton. Evol. Dev. 3, 1–15.

Baum, D.A., 1998. The evolution of plant development. Curr. Opin. Plant Biol. 1, 79–86.

Bender, J., Fink, G., 1998. A *MYB* homolog, *ATR1*, activates tryptophan gene expression in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 95, 5655–5660.

Borevitz, J.O., Xia, Y.J., Dolan L., Schiefelbein, J., 1998. Activation tagging identifies a conserved *MYB* regulator of phenylpropanoid biosynthesis. Plant Cell 12, 2383–2393.

Braun, E.L., Grotewold, E., 1999. Newly discovered plant *c-MYB*-like genes rewrite the evolution of plant myb gene family. Plant Physio. 121, 21–24.

Brubaker, C.L., Wendel, J.F., 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*: Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Amer. J. Bot. 81, 1309–1326.

Brubaker, C.L., Bourland, F.M., Wendel, J.F., 1999. The origin and domestication of cotton. In: Smith WC (ed.) Cotton: Origin, History, Technology, and Production. J. Wiley and Sons, Inc., New York, N.Y.

Brubaker, C.L., Paterson, A.H., Wendel, J.F., 1999. Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome 42, 184–203.

Cronn, R.C., Small, R.L., Wendel, J.F., 1999. Duplicated genes evolve independently following polyploid formation in cotton. Proc. Natl. Acad .Sci. USA 96, 14406–14411.

Cronn, R.C., Small, R.L., Haselkorn, T.S., Wendel, J.F., 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. Amer. J. Bot. 89, 707–725.

Cronn, R., Cedroni, M., Grover, C., Haskelkorn, T., Wendel, J.F., 2002. PCR-mediated recombination in a polyploid plant. Theor. Appl. Genet. 104, 482–489

Cubas, P., Coen, E., Martinez-Zapater, J.M., 2001. Ancient asymmetries in the evolution of flowers. Curr. Biol. 11, 1050–1052.

Doebley, J., Lukens, L., 1998. Transcriptional regulators and the evolution of plant form. Plant Cell 10, 1075–1082.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151, 1531–1545.

Fryxell, P.A., 1979. The natural history of the cotton tribe. Texas A&M University Press, College Station, T.X.

Gocal, G.F.W., Sheldon, C.C., Gubler, F., Moritz, T., Bagnall, D.J., MacMillan, C.P.,.Li, S.F., Parish, R.W., Dennis, E.S., Wiegel, D., King, R.W., 2001. *GAMYB-like* genes, flowering, and gibberellin signaling in *Arabidopsis*. Plant Physiol. 127, 1682–1693.

Hurst, L.D., Pal, C., 2001. Evidence for purifying selection acting on silent sites in BRCA1. Trends Genet. 17, 62–65.

Jin, H., Martin, C., 1999. Multifunctionality and diversity within the plant MYB gene family. Plant Mol. Biol. 41, 577–585.

Kirik, V., Kolle, K., Wohlfarth, T., Misera, S., Baumlein, H., 1998. Ectopic expression of a novel MYB gene modifies the architecture of the *Arabidopsis* inflorescence. Plant J. 13, 729–742.

Kranz, H., Denekamp, M., Greco, R., Jin, H., Leyva, A., Meissner, R.C., Petroni, K., Urzainqui, A., Bevan, M., Martin, C., Smeekens, S., Tonelli, C., Paz-Ares, J., Weisshaar, B., 1998. Towards functional characterisation of the members of the *R2R3-MYB* gene family from *Arabidopsis thaliana*. Plant J. 16, 263–276.

Kranz, H., Scholz, K., Weisshaar, B., 2000. c-Myb oncogene-like genes encoding three MYB repeats occur in all major plant lineages. Plant J. 21, 231–235.

Loguercio, L.L., Zhang, J.Q., Wilkins, T.A., 1999. Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.). Mol. Gen. Genet. 261, 660–671.

Luo, D., Carpenter, R., Copsey, L., Vincent, C., Clark, J., Coen, E., 1999. Control of organ asymmetry in flowers of *Antirrhinum*. Cell 99, 367–376.

Lynch, M. and Force, A., 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154, 459–473.

Martin, C. and Paz-Ares, J., 1997. MYB transcription factors in plants. Trends Genet. 13, 67–73.

Meissner, R.C., Hailing, J., Cominelli, E., Denekamp, M., Fuertes, A., Greco, R., Kranz, H.D., Penfield, S., Petroni, K., Urzainqui, A., Martin, C., Paz-Arez, J., Smeekens, S., Tonelli, C., Weisshaar, B., Baumann, E., Klimyuk, V., Marillonnet, S., Patel, K., Speulman, E., Tissier, A.F., Bouchez, D., Jones, J.J.D., Pereira, A., Weisman, E., Bevan, M., 1999. Function search in a large transcription factor gene family in *Arabidopsis*: Assessing the

potential of reverse genetics to identify insertional mutations in R2R3 MYB genes. Plant Cell 11, 1827–1840.

Noda, K,, Glover, B.J., Linstead, P., Martin, C., 1994. Flower colour intensity depends on specialized cell shape controlled by a Myb-related transcription factor. Nature 369, 661–664.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., Sekiya, T., 1989. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA 86, 2766–2770.

Payne, T., Clement, J., Arnold, D., Lloyd, A., 1999. Heterologous myb genes distinct from GL1 enhance trichome production when overexpressed in *Nicotiana tabacum*. Development 126, 671–682.

Penfield, S., Meissner, R.C., Shoue, D.A., Carpita, N.C., Bevan, M.W., 2001. *MYB61* is required for mucilage deposition and extrusion in the *Arabidopsis* seed coat. Plant Cell 13, 2777–2791.

Purugganan, M.D., 1998 The molecular evolution of development. BioEssays 20, 700–711.

Rabinowicz, P.D., Braun, E.L., Wolfe, A.D., Bowen, B., Grotewold, E., 1999. Maize *R2R3 MYB* genes: sequence analysis reveals amplification in the higher plants. Genetics 153, 427–444.

Reichmann, J.L., Heard, J., Martin, G., Reuber, L., Jiang, C.Z., Keddie, J., Adam, L., Pineda, O., Ratcliffe, O.J., Samaha, R.R., Creelman, R., Pilgrim, M., Broun, P., Zhang, J.Z., Ghandehari, D., Sherman, B.K., Yu, G.L., 2000. *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes. Science 290, 2105–2110.

Reinisch, A.R., Dong, J.M., Brubaker, C.L., Stelly, D.M., Wendel, J.F., Paterson, A.H., 1994. A detailed RFLP map of cotton, *Gossypium hirsutum* × *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138, 829–847.

Romero, I., Fuertes, A., Benito, B.J., Malpica, J.M., Leyva, A., Paz-Ares, J., 1998. More than 80 R2R3-MYB regulatory genes in the genome of *Arabidopsis thaliana*. Plant J. 14, 273–284.

Rozas, J., Rozas, R., 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15, 174–175.

Seelanan, T., Schnabel, A., Wendel, J.F., 1997. Congruence and consensus in the cotton tribe. Systematic Botany 22, 259–290.

Slabaugh, M.B., Huestis, G.M., Leonard, J., Holloway, J.L., Rosato, C., Hongtrakul, V., Martini, N., Toepfer, R., Voetz, M., Schell, J., Knapp, S.J., 1997. Sequence-based genetic markers for genes and gene families: single-stranded conformational polymorphisms for the fatty acid synthesis genes of *Cuphea*. Theor. Appl. Genet. 94, 400–408.

Small, R.L., Wendel, J.F., 2000. Copy number lability and evolutionary dynamics of the *Adh* gene family in diploid and tetraploid cotton (*Gossypium*). Genetics 155, 1913–1926.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T., Wendel, J.F., 1998. The tortoise and the hare: choosing between noncoding plastome and nuclear *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. Amer. J. Bot. 85, 1301–1315.

Song, K., Osborn, T.C., 1994. A method for examining expression of homologous genes in plant polyploids. Plant Mol. Biol. 26, 1065–1071.

Stracke, R., Weber, M. Weisshaar, B., 2001. The R2R3-MYB gene family in *Arabidopsis thaliana*. Curr. Opin. P. Biol. 4, 447–456.

Swofford, D.L, 1998. PAUP*: phylogenetic anaylsis using parsimony (* and other methods), Version 4.0b1. Sunderland, MA: Sinauer Associates, Inc.

Tajima, F., 1993. Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135, 599–607.

Wada, T., Tachibana, T., Shimura, Y., Okada, K., 1997. Epidermal cell differentiation in *Arabidopsis* determined by a *MYB* homolog, CPC. Science 277, 1113–1116.

Wendel, J.F., 1989 New World tetraploid cottons contain Old World cytoplasm. Proc. Acad. Natl. Sci. USA 86, 4132–4136.

Wendel, J.F., 2000. Genome evolution in polyploids. Pl. Mol. Biol. 42, 225–249.

Wendel, J.F., Olson, P.D., Stewart, JMcD, 1989. Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. Amer. J. Bot. 76, 795–806.

Wendel, J.F., 1995. Cotton. In: N. Simmonds and J. Smartt (eds), Evolution of Crop Plants. Longman, London pp 358–366.

Wendel, J.F., Schnabel, A., Seelanan, T., 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc. Natl. Acad. Sci. USA 92, 280–284.

Wilkins, T.A., Jernstedt, J.A., 1999. Molecular genetics of developing cotton fibers. In: Basra AS (ed.), Cotton Fibers. Haworth Press, New York pp. 231–267.

Wilkins, T.A., Smart, L.B., 1996. Isolation of RNA from plant tissue. In: A Laboratory Guide to RNA. Krieg PA (ed.) Wiley-Liss, Inc. pp. 21–40.

Yip, S.P., Hopkinson, D.A., Whitehouse, D.B., 1999. Improvement of SSCP analysis by use of denaturants. Biotechniques 27, 20–24.

Zhang. P., Chopra, S., Peterson, T., 2000. A segmental gene duplication generated differentially expressed *myb*-homologous genes in maize. Plant Cell 12, 2311–2322.