

Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification, phylogenomic characterization, chromosome localization, and linkage mapping

Chuanfu An · Sukumar Saha · Johnie N. Jenkins ·
Din-Pow Ma · Brian E. Scheffler · Russell J. Kohel ·
John Z. Yu · David M. Stelly

Received: 6 August 2007 / Accepted: 11 February 2008 / Published online: 13 March 2008
© Springer-Verlag 2008

Abstract R2R3-MYB transcription factors of plants are involved in the regulation of trichome length and density. Several of them are differentially expressed during initiation and elongation of cotton fibers. We report sequence phylogenomic characterization of the six *MYB* genes, their chromosomal localization, and linkage mapping via SNP marker in AD-genome cotton ($2n = 52$). Phylogenetic grouping and comparison to At- and Dt-genome putative ancestral diploid species of allotetraploid cotton facilitated differentiation between genome-specific polymorphisms (GSPs) and marker-suitable locus-specific polymorphisms (LSPs). The

SNP frequency averaged one per 77 bases overall, and one per 106 and 30 bases in coding and non-coding regions, respectively. SNP-based multivariate relationships conformed to independent evolution of the six *MYB* homoeologous loci in the four tetraploid species. Nucleotide diversity analysis indicated that the six *MYB* loci evolved more quickly in the Dt- than At-genome. The greater variation in the Dt-D genome comparisons than that in At-A genome comparisons showed no significant bias among synonymous substitution, non-synonymous substitution, and nucleotide change in non-coding regions. SNPs were concordantly mapped by deletion analysis and linkage mapping, which confirmed their value as candidate gene markers and indicated the reliability of the SNP discovery strategy in tetraploid cotton species. We consider that these SNPs may be useful for genetic dissection of economically important fiber and yield traits because of the role of these genes in fiber development.

Communicated by D. Lightfoot.

Nucleotide sequence data reported are available in GenBank database under the accession numbers EU249397 to EU249456.

Disclaimer: Mention of trade names or commercial products does not constitute a guarantee or warranty of the product by US Department of Agriculture and does not imply its approval to the exclusion of other products that may also be suitable.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0732-4) contains supplementary material, which is available to authorized users.

C. An
Department of Plant and Soil Sciences,
Mississippi State University, Mississippi State, MS 39762, USA

S. Saha (✉) · J. N. Jenkins
USDA-ARS, Crop Science Research Laboratory,
P.O. Box 5367, Mississippi State, MS 39762, USA
e-mail: Sukumar.Saha@ars.usda.gov;
ssaha@msa-msstate.ars.usda.gov

D.-P. Ma
Department of Biochemistry and Molecular Biology,
Mississippi State University, Mississippi State, MS 39762, USA

Introduction

Cotton (*Gossypium* spp.) is the world's most important natural textile fiber warranting increased exploration of

B. E. Scheffler
USDA-ARS-CGRU, Stoneville, MS 38776, USA

R. J. Kohel · J. Z. Yu
USDA-ARS, Crop Germplasm Research Unit,
2765 F&B Road, College Station, TX 77845, USA

D. M. Stelly
Department of Soil and Crop Sciences,
Texas A&M University, College Station,
TX 77843-2474, USA

fiber-related traits through various molecular genetic approaches. Currently, two types of molecular markers are primarily used in molecular mapping of cotton genome. One is genomic markers which primarily target non-coding regions such as RFLP (Reinisch et al. 1994), RAPD (Kohel et al. 2001), AFLP (Mei et al. 2004), STS (Rong et al. 2004), and SSR (Zhang et al. 2002; Frelichowski et al. 2006). The other is candidate gene markers represented by EST–SSR (Chee et al. 2004; Park et al. 2005; Guo et al. 2007), cDNA probe-based STS or RFLP markers (Rong et al. 2004), and SNP i.e. single nucleotide polymorphism (An et al. 2007). Development of candidate gene markers has received much attention in recent years because of the possible association of functional genes with complex traits. However, the low polymorphism level of cDNA probe-based STS or RFLP markers hampered candidate gene mapping (Rong et al. 2004). SNPs have recently been used as the choice for candidate gene markers in many plant species and are reported to be the most abundant molecular markers (Cho et al. 1999; Ching et al. 2002; Zhang et al. 2003; Zhu et al. 2003). However, SNP development in cotton is impeded by its allotetraploid nature, high repetitive DNA content, and inadequate genome sequence information.

The candidate gene approach is widely accepted as a strategy for identification of loci influencing complex and economically important traits (Faris et al. 1999; Giroux et al. 2000; Pflieger et al. 2001; Beecher et al. 2002). Candidate gene markers derived from resistance genes or deference response genes were placed on regions containing major resistance QTL in wheat (Faris et al. 1999), pepper (Pflieger et al. 1999), and rice (Wang et al. 2001). The storage protein genes for puroindoline in wheat (Giroux et al. 2000) and hordoindolines in barley (Beecher et al. 2002) were both implicated to play a role in grain hardness and texture by QTL analysis. Markers developed from genes related to carbohydrate and nitrogen metabolism were found to be associated with sugar content and yield in sugar beet (Schneider et al. 2002). Wilson et al. (2004) detected significant association between candidate genes involved in kernel starch biosynthesis and traits for maize kernel composition and starch quality. In cotton, Rong et al. (2007) also found evidence of a general association between concentrations of candidate genes and cotton fiber-related QTL.

R2R3-MYB transcription factors, characterized by two imperfect repeats (R2 and R3) in the DNA-binding domain, are one of the largest regulatory gene families in plants (Riechmann et al. 2000). Some of them were shown to control trichome initiation, expansion, branching, and maturation in *Arabidopsis* (Oppenheimer et al. 1991; Glover et al. 1998; Szymanski et al. 2000; Schiefelbein 2003). Cotton fibers are elongated trichomes derived from ovule epidermis. Previous reports suggested a similarity in genetic con-

trol of MYB transcription factors in *Arabidopsis* trichomes and cotton fibers (Suo et al. 2003; Wang et al. 2004; Humphries et al. 2005; Perez-Rodriguez et al. 2005; Wu et al. 2006). Expression analysis demonstrated that six R2R3-MYB transcription factors were expressed in fiber cells but regulated differentially during fiber initiation and expansion (Loguercio et al. 1999; Cedroni et al. 2003). In addition, several other MYB genes have been indicated to play an important role in cotton fiber initiation (Suo et al. 2003; Hsu et al. 2005; Lee et al. 2006; Yang et al. 2006).

Here we report the sequence phylogenomic characterization of the six MYB genes in selected tetraploid and diploid cotton species, their chromosomal locations and molecular linkage map using candidate gene derived SNP markers. The chromosomal locations and genetic linkage mapping of SNP markers with framework SSR markers will improve the resolution of the cotton comparative map. SNP markers derived from MYB genes in this study will be useful as diagnostic markers for exploration of the roles of these candidate genes in complex fiber traits.

Materials and methods

Plant materials

HS46 and MARCABUCAG8US-1-88 (MAR), two *G. hirsutum* (AD₁) lines of diverse agronomic and fiber properties, and three lines of other tetraploid species including *G. barbadense* L. (AD₂, accession 3-79), *G. tomentosum* Nuttall ex Seemann (AD₃), and *G. mustelinum* Miers ex Watt (AD₄) were used for PCR amplification, cloning, and sequencing of the six MYB genes. Chromosomal assignment of SNP markers was accomplished using three different sets of hypoaneuploid F₁ stocks developed from an interspecific cross between TM-1 (genetic standard for *G. hirsutum*, AD₁) and one of the three species, 3-79, *G. tomentosum* or *G. mustelinum*, together with one set of euploid interspecific backcrossed chromosome substitution lines (CS-B, BC₅S₁) of 3-79 in TM-1. Hypoaneuploid F₁ cytogenetic stocks between TM-1 and 3-79 consisted of 10 primary monosomic and 28 monotelodisomic lines; whereas, hypoaneuploid F₁ lines between TM-1 and *G. tomentosum* included 11 primary monosomic and 27 monotelodisomic lines (Liu et al. 2000; Saha et al. 2006b). The new hypoaneuploid F₁ chromosome substitution stocks between TM-1 and *G. mustelinum* (unpublished information) were also used for deletion analysis. Euploid CS-B stocks contain 12 different chromosome and 8 chromosome arm substituted from 3-79 in TM-1 background (Stelly et al. 2005). Fresh leaves were collected from individual plant, frozen in liquid nitrogen, and then subjected to genomic DNA extractions by a Qiagen DNeasy plant maxi kit

(Qiagen Inc., Valencia, CA, USA). A set of 186 recombinant inbred lines (RILs) generated from an interspecific cross between TM-1 and 3-79 were used as a mapping population for constructing molecular linkage map of SNP markers specific to the *MYB* genes and the selected framework SSR markers in cotton (Park et al. 2005; Frelichowski et al. 2006).

PCR amplification, cloning, and sequencing

Gene-specific PCR primers of *MYB1* (COT105 and COT106), *MYB2* (Myb2F and COT108), *MYB3* (Myb3F and COT110), *MYB4* D-genome locus (COT111 and COT112), and *MYB6* (Myb6F and COT116) were adopted from Loguercio et al. (1999). Gene-specific PCR primers of *MYB4* A-genome locus (Myb4A_F and Myb4A_R) and *MYB5* (Myb5_F and Myb5_R) were designed based on GenBank deposited sequences generated from the previous works by Loguercio et al. (1999) and Cedroni et al. (2003) (Table 1). Pfu polymerase (Stratagene, La Jolla, CA, USA) was used for PCR amplification following the protocol described elsewhere (An et al. 2007). The PCR products were separated on a 1% (w/v) agarose gel and purified using QIAEX II gel extraction kit (Qiagen Inc, Valencia, CA, USA). The purified products were ligated into TOPO TA cloning vector and transformed into TOPO10 competent *E. coli* cells (Invitrogen, Carlsbad, CA, USA). Both strands of the recombinant plasmid were sequenced using an ABI 3730XL automated sequencer with ABI Prism Big-Dye Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems, Foster City, CA, USA). In order to avoid possible complications from PCR recombination (Cronn et al. 2002) and to identify the duplicated copies in the genome, we picked up multiple clones (12 clones) for sequencing of each amplicon and considered one identical sequence from at least three clones.

SNP characterization and phylogenetic analysis

Six *MYB* gene sequences from the five allotetraploid cotton lines together with GenBank deposited sequences from TM-1 and living models of two allotetraploid ancestral genomes: *G. herbaceum* L. (A-genome; accession A1-73) and *G. raimondii* Ulbrich (D-genome; “Galau”) were used for SNP characterization. The GenBank sequences of *Gossypoides kirkii* (Masters) J.B. Hutchinson were used as an outgroup to cotton genus (Malvaceae) in phylogenetic analyses (Cedroni et al. 2003). DNASTAR (DNASTAR Inc., Madison, WI, USA) and Clustalx (Thompson et al. 1997) were used for vector-trimming and sequence alignment. Before SNP characterization, differentiation between paralogous and homoeologous loci was performed by phylogenetic grouping and comparison of sequences from the two

diploid species (An et al. 2007). Phylogenetic analyses were performed by maximum parsimony (MP) method using MEGA 3.1 (Kumar et al. 2004). To determine the confidence levels for each tree, an MP bootstrap analysis with 100 replicates was conducted. DnaSP 4.0 software was used to identify SNP by comparative analysis of aligned sequences from different genotypes at a putative locus (Rozas et al. 2003). Nucleotide diversities (π), haplotype number (H) and diversity (H_d), rate of silent (K_{sil}) and non-synonymous (K_a) substitutions of pairwise comparisons were also calculated by DnaSP 4.0 software (Tajima 1983; Nei 1987; Rozas et al. 2003).

Chromosomal assignment and linkage mapping

In order to minimize the potential problems associated with homoeologous sequences in SNP genotyping, genome-specific (or locus-specific) PCR primers were designed according to sequence differences between two subgenomes in tetraploid cotton when applicable (Table 1). Interspecies SNP primers were designed based on a single nucleotide difference among sequences at a putative locus (each clade or group in the phylogram of individual *MYB* gene) between the genotypes of TM-1 and 3-79, *G. tomentosum* or *G. mustelinum*. The primer was designed to anneal just upstream or downstream of the SNP site as the forward or reverse primer, respectively, so that the polymorphism could be detected by one base extension technology with an ABI Prism SNaPshot™ multiplex kit (Applied Biosystems, Foster City, CA, USA). All the primers used for genotyping are summarized in Table 1. The deletion analysis method frequently used for molecular marker chromosomal assignment in cotton (Liu et al. 2000; An et al. 2007) was employed to assign chromosomal locations for six *MYB* genes using the four sets of cytogenetic stocks mentioned in the plant materials. A total of 90 SSR markers, which are polymorphic between TM-1 and 3-79 and span the cotton genome, were selected based on the information available in cotton microsatellite database (CMD, <http://www.cotton-marker.org/>; Blenda et al. 2006), and used as anchored markers for linkage mapping of sections of selected chromosomes with SNP markers. Chromosomal assignment of the constructed linkage groups was achieved by deletion analysis, comparison to the allele size with CMD panel, published integrated molecular maps (Lacape et al. 2005; Park et al. 2005; Frelichowski et al. 2006; Guo et al. 2007), and the assignment of cotton linkage maps to chromosomes (Wang et al. 2006b). The SSR markers used in this study were fluorescent labeled by Sigma Genosys (The Woodlands, TX, USA) or Applied Biosystems (Foster City, CA, USA). PCR reactions and thermal cycle protocols for genotyping the RILs population were conducted according to the method of Gutierrez et al. (2002). One polymorphic

Table 1 SNP makers derived from six *MYB* genes for genotyping in four allotetraploid cotton species

Gene	Genome	PCR primer (5'-3')	SNP primer (5'-3')	Genotyping result			
				TM-1	3-79	Gm ^a	Gt ^b
<i>MYB1</i>	A	Myb1A_F: GAGTTTCCTCGTATTAATTTTCAGC	Myb1Gbmt_238_R: CTTGTGCAGCTGTTCTTGTGTG	C	T	T	T
		COT106: ACCCTATGAATCCAAGGGTC	Myb1Gt_272_R: CTTGCCCATTTGCTCGGAC	C	C	C	A
		Myb1D_F: GGAAACAAGTGAAGTTTCTCCTTTCG COT106: ACCCTATGAATCCAAGGGTC	Myb1Gb_500_F: CTGTCAAAAATTCAAAATTCAAAAT	G	T	G	G
<i>MYB2</i>	A	Myb2A_F: CTGTTCTGACATCATGGTTTATT	Myb2Gb_204_R: TAGTTCAACCATCTCAACC	T	A	T	T
		Myb2_R: GGCTAATAATGGCTCCAAAAGAAGG	Myb2Gb_334_F: TGATTGCTGGGAGACTTC	C	T	C	C
		Myb2D_F: GTTCTGTTCTGACATCATGGTTTATC Myb2_R: GGCTAATAATGGCTCCAAAAGAAGG	N/A ^c	N/A	N/A	N/A	N/A
<i>MYB3</i>	A/D ^d	Myb3_F: GGGCCACTAAA GAATGGAGCA	Myb3Gt_730_F: GAGCTTTTGTGTTGTTTATCTTCT	A	A	A	G
		COT110: GCTACAGTTCACTATGTCGG	Myb3Gb_742_R: CATGGATTGCATCAAAATTCAT	G	A, G ^e	G	G
		Myb4A_F: TAACTCAAAGCATGCCAGTC Myb4A_R: CACTAAAGCAGAAAGGAGCATTAG	Myb4Gbmt_428_F: CTGAGAAAAGGGACTGGTAT	C	T	T	T
<i>MYB5</i>	A/D ^d	COT111: CCATTAACCTCAAAGCATGCC	Myb4Gbmt_73_R: CCTTCATCGACCCCTTCGCTA	G	T	T	T
		COT112: CAGGAGGAACAAGGAGGAGC	Myb4Gbmt_105_R: TCCAGTATCTTTCTGGGTACTGC	A	G	G	G
		Myb5_F: GACCTTTTGAACCTGTGAAGGT Myb5_R: CTGCTGTAAAAACCAACATTC	Myb5Gbmt_511_R: ACCAACCAACAACCAACCAATTC	C, T ^f	T	T	T
<i>MYB6</i>	A	Myb6A_F: CGTTCCTCCCTCCCAAAAGCT COT116: ACAATGGCGTGCATGTTGCC	Myb6Gm_341_F: CATTTCCAAAAACATAACGTTTAC	G	G	A	G
		Myb6A_R: CGTTCCTCCCTCCCAAAAGCC COT116: ACAATGGCGTGCATGTTGCC	Myb6Gb_856_F: GTAACAATAATGATGAAGAAGAAC	A	G	A	A

^a Abbreviation of *G. mustelinum*^b Abbreviation of *G. tomentosum*^c No suitable SNP marker could be designed in D-genome sequence of gene *MYB2*^d Not applicable for genome-specific PCR primers design^e 3-79 gave two alleles for this SNP marker, one from each subgenome. We utilized CS-B lines for allele "A" (in D-genome) chromosomal assignment^f TM-1 gave two alleles for this SNP marker, one from each subgenome. Allele "C" was chromosomally assigned by deletion analysis with hypoaneuploid F₁ cytogenetic stocks

SNP marker between TM-1 and 3-79 was selected, if available, from each gene for linkage mapping. The procedures of SNP marker genotyping described in An et al. (2007) were employed for cytogenetic stock and RIL population genotyping. An automated capillary electrophoresis system ABI3100 Genetic Analyzer with GeneMapper software 4.0 (Applied Biosystems, Foster City, CA, USA) was used to analyze both PCR-amplified DNA fragments of SSR markers and the single nucleotide extension of SNP markers. The genotyping output data of both SNP and SSR markers were coded for linkage analysis using JoinMap[®] 4 (Van Ooijen 2006). The fit of marker segregation to the 1:1 ratio expected was evaluated according to Chi-square tests. Recombination frequencies were converted into map distances (centiMorgan, cM) using the Kosambi mapping function (Kosambi 1944) and linkage groups were determined at LOD scores ≥ 6 .

Results

SNP characterization and haplotype analyses of six *MYB* genes

In vitro SNP discovery through amplicon cloning and sequencing was accomplished by homoeologous differentiation and gene specific fragment amplification in cotton (Supplementary Fig. 1; Table 2). In this study, no dupli-

cated or heterogeneous loci were found within each subgenome. SNPs and indels were detected from 8,301 bp of aligned sequences (7,084 and 1,217 bp of coding and non-coding regions, respectively). From the eight cotton genotypes, 108 SNPs were detected from *Gossypium* species (Table 2), giving an average SNP frequency of one SNP every 77 bases. Results showed the presence of one SNP per 106 bp in the coding regions and one SNP per 30 bp in the non-coding regions (Table 2). The SNP distribution varied among the six examined genes. The highest rate of SNP occurrence was observed in *MYB6* (one SNP every 34 bp) and the lowest rate of SNP frequency was present in *MYB3* (one SNP every 260 bp). Transitions (“A/G” or “C/T”) were the most common cause of sequences variation in the selected cotton genotypes (49%) compared to transversions (“A/T”, “G/C”, “A/C” or “G/T”, 26%) and indels (25%). In *MYB6*, two nucleotide (“C” and “T”) substitutions were observed in three indel positions (A-genome sites 101 and 111, D-genome site 99). A significant bias to “T” insertion/deletion was detected in the overall sequences (59.30%). In coding regions of the six *MYB* genes, 41 out of 67 cSNPs (SNPs in coding region) sites were predicted to result in amino acid changes (Table 2). The number of haplotypes defined by sequence polymorphism ranged from two to seven among the seven selected cotton genotypes, and haplotype diversity varied from 0.286 ± 0.196 to 1.000 ± 0.076 among six *MYB* genes (Table 2 and Supplementary Tables 1–12).

Table 2 SNP characterization of the six *MYB* genes in selected cotton genotypes

Gene	Genome	Haplotypes ^a		Transitions		Transversions			Indels				Coding region		Total		
		H	H _d	A/G	T/C	A/T	G/C	A/C	G/T	A	C	G	T	Length (bp)	SNP number (non-synonymous changes)	Length (bp)	SNP number
<i>MYB1</i>	A	5	0.857 ± 0.137	2	2	1	0	0	1	0	0	0	0	492	6 (4)	492	6
	D	4	0.714 ± 0.181	2	2	0	0	0	0	0	0	0	0	492	4 (3)	492	4
<i>MYB2</i>	A	4	0.714 ± 0.181	1	2	1	1	1	1	0	0	0	2	473	5 (3)	551	9
	D	4	0.714 ± 0.181	1	3	1	0	0	0	0	0	0	1	473	5 (4)	550	6
<i>MYB3</i>	A	2	0.286 ± 0.196	0	0	0	1	0	0	0	0	0	0	537	1 (0)	778	1
	D	4	0.714 ± 0.181	3	1	0	1	0	0	0	0	0	0	537	2 (1)	779	5
<i>MYB4</i>	A	5	0.857 ± 0.137	1	3	2	2	2	2	0	0	0	0	842	12 (8)	853	12
	D	5	0.857 ± 0.137	2	3	2	0	1	0	0	0	0	0	842	8 (5)	853	8
<i>MYB5</i>	A	3	0.714 ± 0.127	1	0	1	0	0	0	0	0	0	0	571	2 (1)	591	2
	D	2	0.286 ± 0.196	1	2	0	0	0	0	0	0	0	0	571	3 (0)	591	3
<i>MYB6</i>	A	5	0.857 ± 0.137	2	6	1	2	0	0	0	4 ^b	1	5	627	7 (4)	882	21
	D	7	1.000 ± 0.076	6	7	2	0	2	0	1	4 ^c	1	8	627	12 (8)	889	31

^a Sites with alignment gaps were not considered for haplotype analysis; H haplotype number; H_d haplotype diversity

^b At positions 101 and 111 of the aligned sequences, two kinds of nucleotide substitution (“C” and “T”) occur at the indel positions. Here, we considered them as “C” nucleotide indel in data analysis

^c At position 99 of the aligned sequences, two kinds of nucleotide substitution (“C” and “T”) occurs at the indel position. Here, we also considered it as “C” nucleotide indel in data analysis

Phylogenomic sequence characterization

SNP-based multivariate relationships suggested independent evolution of the six *MYB* homoeologous loci in the four tetraploid species. Parsimony analyses revealed that sequences (Supplementary Figs. 1, 2) fell into two clades, each containing one of the two homoeologous loci from the allotetraploid cotton lines and the corresponding copy from the progenitor diploid genomes. Pairwise comparisons of the nucleotide diversity (π) of the six *MYB* genes in both A- and D-genomes are summarized in Table 3. The π value measures the average number of nucleotide differences per site between two sequences (Nei 1987). The lowest nucleotide diversities occurred among the three *G. hirsutum* lines in both A- and D-genomes. Results from both A- and D-genomes showed the highest nucleotide diversities were between *G. mustelinum* and the extant relatives of the ancestral genome donors. Nucleotide diversities of *MYB* genes were higher in the D–Dt comparisons than for the A–At comparison of the allotetraploid cotton species, indicating that *G. herbaceum* may be a closer ancestor of the At-genome donor than *G. raimondii* is of the Dt-genome donor.

To further explore the nature of substitutions contributing to overall divergence in cotton, pairwise comparisons among orthologous copies for the six *MYB* genes of both A- and D-genomes are tabulated separately for non-synonymous substitution (K_a), silent substitution (K_{sil}), and the $K_a:K_{sil}$ ratio (Table 4). K_a and K_{sil} values in the D–Dt comparisons were higher than the corresponding values in the A–At comparisons, except for the comparison between MAR (*G. hirsutum*) and its two genome living models of K_{sil} value. Contributing to the relatively level of D–Dt differentiation were greater amino acid substitutions, nucleotide changes in non-coding regions, and synonymous changes in the coding regions. Although these predictions were based on the genomic sequence, they may allow speculation of evolutionary constraints placed on amino acid substitutions without knowing the exact effect of the SNPs on predicted codons. Nucleotide diversities among the three

G. hirsutum lines in the Dt-genome were higher than those in the At-genome, indicating six *MYB* genes loci in Upland cotton Dt-genome exhibited a faster evolutionary rate than the At-genome (Table 3). Most of the substitution ratios ($K_a:K_{sil}$) of pairwise comparisons were less than 1, indicating the possibility of a high level of evolutionary constraint placed on amino acid substitution in the six *MYB* genes (Table 4).

Chromosome localization of six *MYB* genes

Hypoaneuploid stocks, developed from three interspecific crosses between TM-1 (*G. hirsutum*) and 3-79 (*G. barbadense*), *G. tomentosum* or *G. mustelinum*, and one set of euploid interspecific backcrossed chromosome substitution lines (CS-B, BC₅S₁) of 3-79 in TM-1 were used for chromosomal assignment of SNP markers by deletion analysis (An et al. 2007). Thirteen different SNP sites between the common parent TM-1 and 3-79, *G. mustelinum* or *G. tomentosum*, respectively, were selected for SNP primer design in six *MYB* genes (Table 1). We confirmed our identification of chromosomal locations using deletion lines from different sources. Due to the conserved character of the homoeologous sequences in the genes *MYB3* and *MYB5*, no suitable genome-specific PCR primers could be designed. However, chromosomal assignment of genome-specific alleles was still possible by euploid CS-B or hypoaneuploid F₁ stocks (Table 1). Moreover, no SNP marker could be designed from the Dt-genome of gene *MYB2*. Therefore, only the At-genome location was considered for chromosomal assignment by either deletion analysis or linkage mapping. SNP markers used for chromosomal assignment and the according genotyping results are listed in Table 1. Deletion analyses of the six genes were performed using all the available cytogenetic stocks and the results are summarized in Table 5. We detected chromosomal locations of the gene *MYB4* on the long arm of two homoeologous chromosomes: 7 and 16. Only one subgenomic location of gene *MYB1*, *MYB2*, *MYB5*, and *MYB6*

Table 3 Pairwise comparison matrix of the eight cotton lines showing the nucleotide diversity (π value, $\times 10^{-2}$) in A- and D-genomes^a

Genotype ^b	A ₁	TM-1	HS46	MAR	3–79	<i>G. mustelinum</i>	<i>G. tomentosum</i>
D ₅	–	0.5267	0.5752	0.5605	0.5200	0.6823	0.5670
TM-1	0.3703	–	0.1055	0.0905	0.1472	0.2897	0.1748
HS46	0.3705	0	–	0.05267	0.1962	0.3195	0.2050
MAR	0.4085	0.03817	0.0382	–	0.1810	0.2857	0.1898
3–79	0.4180	0.2113	0.2113	0.2495	–	0.3018	0.1875
<i>G. mustelinum</i>	0.5572	0.3310	0.3312	0.2930	0.3795	–	0.2727
<i>G. tomentosum</i>	0.4950	0.2878	0.2690	0.3260	0.3362	0.4558	–

^a The numbers above and below the diagonal line represent sequence comparisons within D- and A-genomes, respectively

^b Taxa *G. herbaceum* and *G. raimondii* are designated by their genome designations, A₁ and D₅, respectively

Table 4 Pairwise comparison matrix of molecular evolutionary rates for the six *MYB* genes in cotton^a

Genotype ^b	A ₁	TM-1	HS46	MAR	3-79	<i>G. mustelinum</i>	<i>G. tomentosum</i>
D ₅		0.0050 ^c	0.0047	0.0051	0.0046	0.0045	0.0040
	–	0.0071	0.0087	0.0071	0.0076	0.0108	0.0094
		0.6956	0.5411	0.7096	0.6009	0.4155	0.4281
TM-1	0.0027		0.0011	0.0015	0.0020	0.0019	0.0014
	0.0067	–	0.0016	0	0.0005	0.0036	0.0023
	0.4040		0.7053	–	4.0690	0.5185	0.6370
HS46	0.0027	0		0.0004	0.0017	0.0017	0.0012
	0.0067	0	–	0.0016	0.0021	0.0052	0.0038
	0.4040	–		0.2211	0.8306	0.3119	0.3087
MAR	0.0032	0	0		0.0021	0.0020	0.0015
	0.0076	0.0009	0.0009	–	0.0005	0.0036	0.0023
	0.4163	0	0		4.2414	0.5463	0.6815
3-79	0.0039	0.0020	0.0020	0.0020		0.0015	0.0010
	0.0051	0.0034	0.0034	0.0042	–	0.0041	0.0027
	0.7647	0.5842	0.5842	0.4646		0.3648	0.3804
<i>G. mustelinum</i>	0.0038	0.0018	0.0018	0.0018	0.0030		0.0009
	0.0084	0.0066	0.0066	0.0058	0.0051	–	0.0050
	0.4484	0.2764	0.2764	0.3179	0.6007		0.1873
<i>G. tomentosum</i>	0.0039	0.0015	0.0015	0.0015	0.0027	0.0027	
	0.0078	0.0061	0.0061	0.0070	0.0045	0.0081	–
	0.4989	0.2452	0.2452	0.2143	0.6022	0.3176	

^a The numbers above and below the diagonal line represent D- and A-genome sequences comparisons, respectively

^b Taxa *G. herbaceum* and *G. raimondii* are designated by their genome designation, A₁ and D₅, respectively

^c Three parameters were used to characterize each comparison: nonsynonymous substitution per nonsynonymous sites in coding sequences (K_a , top), substitutions per site including intron and synonymous sites (K_{sil} , middle), and $K_a:K_{sil}$ ratio (bottom)

was found by deletion analysis using SNP markers, which was on the long arm of chromosome 18, short arm of chromosome 8, short arm of chromosome 11, and short arm of chromosome 11, respectively. We do not have complete coverage for all the chromosomes in the cytogenetic stocks. The putative chromosome location of gene *MYB3* in Dt-genome could not be determined due to incomplete coverage of Dt-genome; however, it is probably on one of the chromosomes for which we do not have aneuploid stock coverage (long arm of chromosomes 14, 15, or chromosomes 19, 21, 23, and 24).

Linkage mapping of *MYB* genes by SNP markers

Framework SSR markers were utilized to construct linkage maps with SNP markers. We used 186 RILs, from the cross of TM-1 and 3-79, for genotyping by 90 SSR markers and five polymorphic SNP markers specific to genes *MYB1*, *MYB2*, *MYB4*, and *MYB6*. Genetic linkage mapping results confirmed the deletion analysis for the chromosomal locations of *MYB1*, *MYB2*, and *MYB4*. Linkage mapping also revealed chromosomal locations of two genes' homoeologous loci (At-genome of gene *MYB1* and Dt-genome of

gene *MYB6*), which were on chromosome 13 and 21, respectively (Table 5). Moreover, it showed the linkage relationship between 15 SSR markers and 5 SNP markers (Fig. 1). Three SNP markers showed distorted segregation in the mapping population. The segregation of SNP markers Myb1Gbmt_238_R and Myb4Gbmt_105_R was skewed toward TM-1 and the segregation of SNP marker Myb2Gb_204_R was skewed toward 3-79.

Discussion

SNP in cotton

Efficient SNP discovery in polyploids, such as cotton, must address the problem with appropriate methods of distinguishing between genome-specific polymorphisms (GSPs) and locus-specific polymorphisms (LSPs). In this study, we reduced the possibility of identifying false SNP by applying the following approaches: (1) designing PCR primers from well-characterized genes to generate an amplicon pool from each genotype; (2) sequencing multiple clones to avoid random sequencing errors and to ensure getting the duplicated

Table 5 Chromosomal locations of the six *MYB* genes with previously reported QTL

Gene	Subgenome	Chromosomal location		Previously reported QTL associated with specific chromosome or chromosome arm ^a
		Deletion analysis	Linkage mapping	
<i>MYB1</i>	A	–	13	EL ^{b, m, n} ; FF ⁿ ; FL ^{c, k} ; FS ^{k, n} ;
	D	18Lo	18	FS ^{e, m} ; EL ^{b, e} ; FF ^{d, k} ; FL ^{j, u} ; 2.5% SL ^e ; 50% SL ^e ; FU ^u ; LP ^u
<i>MYB2</i>	A	8sh	8	EL ^{i, n, q} ; FL ^{c, h, n} ; FU ^u ; LI ^h ; LY ^{g, h} ; SCY ⁱ ; MIC ^l ; FS ^{h, l, n, q, u}
	D	–	–	–
<i>MYB3</i>	A	–	–	–
	D	Possibly 14Lo/15Lo/19/21/23/24	–	EL ^{b, e, i, k, m, n, q, t} ; FF ^{d, k, n} ; FL ^{c, h, i, k, l, n, q, r, s, t} ; LP ^s ; FU ^{j, u} ; LY ^{g, h, r, s} ; MIC ^{h, l, q, r, s, t} ; SCY ^{h, p, r, s} ; FS ^{e, h, i, j, k, m, l, n, q, s, t, u}
<i>MYB4</i>	A	7Lo	–	EL ^q ; FL ^{c, q} ; FU ^p ; SCY ^{g, h}
	D	16Lo	16	EL ^o ; FF ^{j, k} ; FL ^{o, q} ; LI ^o ; LP ^o ; MIC ^q ; FS ^{k, q, r}
<i>MYB5</i>	A	11sh	–	EL ^{b, r, s} ; FF ^d ; FL ^{c, n} ; FU ⁿ ; LP ^f ; LY ^{g, h} ; MIC ^{l, q} ; FS ^{n, q}
	D	–	–	–
<i>MYB6</i>	A	11sh	–	EL ^{b, r, s} ; FF ^d ; FL ^{c, n} ; FU ⁿ ; LP ^f ; LY ^{g, h} ; MIC ^{l, q} ; FS ^{n, q}
	D	–	21	EL ^b ; FF ^{d, n} ; FL ^{c, q, r} ; FS ^{i, n, p}

^a EL Elongation; FF fineness; FL length; FS strength; FU uniformity; LI lint index; LP lint percentage; LY lint yield; MIC micronaire; SCY seed cotton yield; 2.5% SL length at 2.5%; 50% SL length at 50%, ^bChee et al. (2005a), ^cChee et al. (2005b), ^dDraye et al. (2005), ^eFrelichowski et al. (2006), ^fGuo et al. (2006), ^gHe et al. (2005), ^hHe et al. (2007), ⁱJiang et al. (1998), ^jKohel et al. (2001), ^kLacape et al. (2005), ^lLin et al. (2005), ^mPark et al. (2005), ⁿPaterson et al. (2003), ^oRen et al. (2002), ^pSaranga et al. (2001), ^qShen et al. (2005), ^rShen et al. (2006), ^sShen et al. (2007), ^tWang et al. (2006a), ^uZhang et al. (2005)

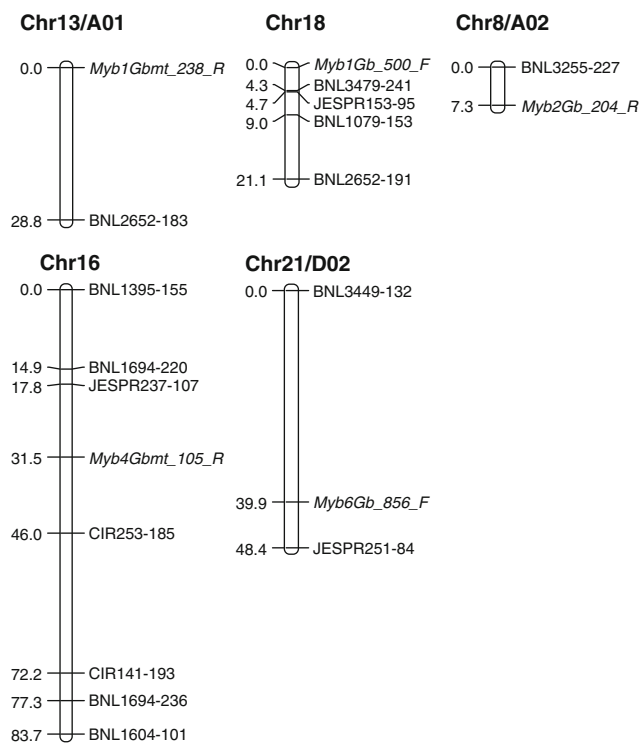


Fig. 1 Linkage maps of selected SNP markers derived from *MYB* genes with SSR markers. The names of the linkage groups A01, A02, and D02 and chromosomal locations were as per Wang et al. (2006b)

loci of the gene; (3) putative locus identification by phylogenetic clustering and comparing to the two progenitor diploid genome species of allotetraploid cottons; (4) designing

locus-specific PCR and SNP primer for SNP marker genotyping to confirm the reliability of the procedures (An et al. 2007). Thus, a total of 108 putative SNPs were identified among selected genotypes at the same locus. The average frequency was one SNP per 77 bp (1.30%), with one SNP per 106 bp (0.94%) and one SNP per 30 bp (3.33%) in coding and non-coding regions, respectively. In *Arabidopsis thaliana*, the rate of variation per nucleotide were detected as 1.09 and 0.27% in *GLI* gene (a member of the *MYB* gene family) of 26 accessions (Hauser et al. 2001) and *Atmyb2* gene of 20 ecotypes (Kamiya et al. 2002), respectively. In cotton, the average rate of SNP per nucleotide was observed as 2.35% in six *EXPANSIN A* genes (An et al. 2007). Another pilot SNP study revealed the rate of variation per nucleotide was 0.35% between *G. hirsutum* and *G. barbadense* (one SNP every 286 bp), and the variations per nucleotide were 0.14 and 0.37%, respectively within these two species (Rong et al. 2004).

In other crops, Ching et al. (2002) reported the presence of one SNP per 31 bp in non-coding regions and one per 124 bp in coding regions when analyzing 18 maize genes in 36 inbred lines. One SNP in every 273 bp was present in soybean (Zhu et al. 2003). Genome-wide sequence alignment between rice subspecies *Indica* and *Japonica* revealed a polymorphism rate of 1.70 SNP/kb and 0.11 indel/kb (Feltus et al. 2004). In wheat, SNP frequency was one SNP per 540 bp (Somers et al. 2003). The incidence of SNP in barley was reported as one SNP per 27 bases in the intronless *Isa* gene (Bundock et al. 2003), and approximately one

SNP per 131 bases in the exonic region of the *P450* gene family members (Bundock and Henry 2004). Although varying frequencies of SNP per length of DNA sequence have been reported, they are highly dependent upon the kind of sequence data and genotypes used to generate SNP in each species. As expected, we observed more number of SNPs at the interspecific level compared to intraspecific level of six cotton *MYB* genes in this study.

MYB gene phylogenomic features

The cotton genus contains about 50 species with the basic chromosome number of 13. The five tetraploid cotton species (AADD, $2n = 4x = 52$) are a monophyletic assemblage putatively derived from a single allopolyploidization event that occurred 1.5 million years ago (MYA) after divergence of the diploid progenitors about 6.7 MYA (Senchina et al. 2003). The two diploid species that gave rise to the allotetraploids were from the A- and D-genome groups which are best represented by the extant species *G. herbaceum* L. and *G. raimondii* Ulbr., respectively (Wendel and Cronn 2003). Our results showed that the tetraploid *MYB* genes could be broadly separated into two origins representing the putative A- and D-genomes based on their similarity with the sequences of the diploid ancestral species (Supplementary Figs. 1, 2). SNP-based multivariate relationships conformed to independent evolution of the six *MYB* homoeologous loci in the four tetraploid species (Cronn et al. 1999; Cedroni et al. 2003). We observed that the nucleotide diversity was higher in the Dt-genome compared to the At-genome of the three *G. hirsutum* lines. Previous studies with *Adh* (Small et al. 1998, 1999; Small and Wendel 2002) and *FAD2-1* (Liu et al. 2001) showed a faster evolutionary rate in the Dt-genome than in the At-genome of cotton. Reinisch et al. (1994) reported that the RFLP marker polymorphism levels of the Dt-genome were 10% higher than the At-genome. The Dt-genome, from an ancestor that does not produce spinnable fiber, contributes substantially to fiber quality of tetraploid cottons (Jiang et al. 1998; Saranga et al. 2001; Paterson et al. 2003; Lacape et al. 2005; Rong et al. 2007). Many QTLs that positively affect fiber quality have been detected on the Dt-genome (Table 5). In addition, many EST loci associated with fiber development have also been mapped to the Dt-genome (Park et al. 2005). However, some QTL influencing fiber quality and yield have been identified in the At-genome as well (Mei et al. 2004; Frelichowski et al. 2006). Whether the spreading of the At-genome repetitive DNA elements to the Dt-genome (Zhao et al. 1998) or different evolutionary pressures operating on the two genomes (Small and Wendel 2002) caused the different evolutionary dynamics is still obscure. But, all these facts collectively indicated the importance of further

investigations of the Dt-genome for fiber improvement in the tetraploid cottons.

Chromosomal locations of *MYB* genes

The chromosomal locations of six *MYB* genes were identified via deletion analysis or linkage mapping (Table 5; Fig. 1). The low level of polymorphism in molecular markers derived from functional genes such as EST-SSR (Park et al. 2005; Guo et al. 2007) or cDNA probe-based STS or RFLP (Rong et al. 2004) among mapping parents has hindered their use in candidate gene mapping. Results presented here show the great potential for using SNP markers to tag functional genes and improve the comparative maps in cotton.

Previous studies have led to the discoveries of important QTL on different chromosomes in cotton. A comprehensive summary of the previously reported cotton fiber quality and yield component traits related QTL on the same chromosomes as the six *MYB* genes are summarized in Table 5. Analyses on the effects of chromosome-specific introgression in Upland cotton indicated that substitutions for chromosomes 16 and 18 from 3-79 had additive effects related to reduced yield (Saha et al. 2006a). These chromosomes are the locations of *MYB1* and *MYB4* genes. Further studies using topcrosses of 13 CS-B lines with five commercial cultivars showed that chromosomes 7 and 18 (locations of gene *MYB4* and *MYB1*, respectively) had additive effects for fiber strength (Jenkins et al. 2007). Given the role of MYB transcription factors in fiber cell initiation and expansion, the agreement of the chromosomal locations between *MYB* genes and previously reported fiber yield and quality QTL suggested these SNP markers may be useful in studying the association between important fiber development genes and economically important QTL in cotton.

Acknowledgments The authors grateful acknowledge the initial works on *MYB* genes by Drs. T.A. Wilkins, Texas Tech University, TX and J.F. Wendel, Iowa State University, IA. We also thank Drs. I.Y. Abdurakhmonov, Institute of Genetics and Plant Experimental Biology, Academy of Sciences of Uzbekistan, Uzbekistan, J. Li, and J. Wilkinson, Mississippi State University, MS for their valuable suggestions in this paper.

Reference

- An C, Saha S, Jenkins JN, Scheffler BE, Wilkins TA, Stelly DM (2007) Transcriptome profiling, sequence characterization, and SNP-based chromosomal assignment of the *EXPANSIN* genes in cotton. *Mol Genet Genomics* 278:539–553
- Beecher B, Bowman J, Martin JM, Bettge AD, Morris CF, Blake TK, Giroux MJ (2002) Hordoidolines are associated with a major endosperm-texture QTL in barley (*Hordeum vulgare*). *Genome* 45:584–591

- Blenda A, Scheffler J, Scheffler B, Palmer M, Lacape JM, Yu JZ, Jesudurai C, Jung S, Muthukumar S, Yullambalase P, Ficklin S, Staton M, Eshelman R, Ulloa M, Saha S, Burr B, Liu S, Zhang T, Fang D, Pepper A, Kumpatla S, Jacobs J, Tomkins J, Cantrell R, Main D (2006) CMD: a cotton microsatellite database resource for *Gossypium* genomics. *BMC Genomics* 7:132
- Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, haplotype diversity and recombination in the *Isa* gene of barley. *Theor Appl Genet* 109:543–551
- Bundock PC, Christopher JT, Eggler P, Ablett G, Henry RJ, Holton TA (2003) Single nucleotide polymorphisms in cytochrome P450 genes from barley. *Theor Appl Genet* 106:676–682
- Cedroni ML, Cronn RC, Adams KL, Wilkins TA, Wendel JF (2003) Evolution and expression of *MYB* genes in diploid and polyploid cotton. *Plant Mol Biol* 51:313–325
- Chee PW, Rong J, Williams-Coplin D, Schulze SR, Paterson AH (2004) EST derived PCR-based markers for functional gene homologues in cotton. *Genome* 47:449–462
- Chee PW, Draye X, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005a) Molecular dissection of inter-specific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach: I. Fiber elongation. *Theor Appl Genet* 111:757–763
- Chee PW, Draye X, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005b) Molecular dissection of phenotypic variation between *Gossypium hirsutum* and *Gossypium barbadense* (cotton) by a backcross-self approach: III. Fiber length. *Theor Appl Genet* 111:772–781
- Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ (2002) SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics* 3:19
- Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, Dewdney J, Reuber TL, Stammers M, Federspiel N, Theologis A, Yang WH, Hubbell E, Au M, Chung EY, Lashkari D, Lemieux B, Dean C, Lipshutz RJ, Ausubel FM, Davis RW, Oefner PJ (1999) Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat Genet* 23:203–207
- Cronn R, Small RL, Wendel JF (1999) Duplicated genes evolve independently after polyploid formation in cotton. *Proc Natl Acad Sci USA* 96:14406–14411
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor Appl Genet* 104:482–489
- Draye X, Chee PW, Jiang CX, Decanini L, Delmonte TA, Bredhauer R, Smith CW, Paterson AH (2005) Molecular dissection of inter-specific variation between *Gossypium hirsutum* and *G. barbadense* (cotton) by a backcross-self approach: II. Fiber fineness. *Theor Appl Genet* 111:764–771
- Faris JD, Li WL, Liu DJ, Chen PD, Gill BS (1999) Candidate gene analysis of quantitative disease resistance in wheat. *Theor Appl Genet* 98:219–225
- Feltus FA, Wan J, Schulze SR, Estill JC, Jiang N, Paterson AH (2004) An SNP resource for rice genetics and breeding based on subspecies *Indica* and *Japonica* genome alignments. *Genome Res* 14:1812–1819
- Frelichowski JE, Palmer MB, Main D, Tomkins JP, Cantrell RG, Stelly DM, Yu J, Kohel RJ, Ulloa M (2006) Cotton genome mapping with new microsatellites from Acala ‘Maxxa’ BAC-ends. *Mol Genet Genomics* 275:479–491
- Giroux MJ, Talbert L, Habernicht DK, Lanning S, Hemphill A, Martin JM (2000) Association of puroindoline sequence type and grain hardness in hard red spring wheat. *Crop Sci* 40:370–380
- Glover BJ, Perez-Rodriguez M, Martin C (1998) Development of several epidermal cell types can be specified by the same MYB-related plant transcription factor. *Development* 125:3497–3508
- Guo WZ, Ma GJ, Zhu YC, Yi CX, Zhang TZ (2006) Molecular tagging and mapping of quantitative trait loci for lint percentage and morphological marker genes in Upland cotton. *J Integr Plant Biol* 48:320–326
- Guo W, Cai C, Wang C, Han Z, Song X, Wang K, Niu X, Wang C, Lu K, Shi B, Zhang T (2007) A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in *Gossypium*. *Genetics* 176:527–541
- Gutierrez OA, Basu S, Saha S, Jenkins JN, Shoemaker DB, Cheatham CL, McCarty JC (2002) Genetics distance among selected cotton genotypes and its relationship with F_2 performance. *Crop Sci* 42:1841–1847
- Hauser MT, Harr B, Schlotterer C (2001) Trichome distribution in *Arabidopsis thaliana* and its close relative *Arabidopsis lyrata*: molecular analysis of the candidate gene *GLABROUS1*. *Mol Biol Evol* 18:1754–1763
- He DH, Lin ZX, Zhang XL, Nie YC, Guo XP, Feng CD, Stewart JMcD (2005) Mapping QTLs of traits contributing to yield and analysis of genetic effects in tetraploid cotton. *Euphytica* 144:141–149
- He DH, Lin ZX, Zhang XL, Nie YC, Guo XP, Zhang YX, Li W (2007) QTL mapping for economic traits based on a dense genetic map of cotton with PCR-based markers using the interspecific cross of *Gossypium hirsutum* × *Gossypium barbadense*. *Euphytica* 153:181–197
- Hsu CY, Jenkins JN, Saha S, Ma DP (2005) Transcriptional regulation of the lipid transfer protein gene *LTP3* cotton fiber by a novel MYB protein. *Plant Sci* 168:167–181
- Humphries JA, Walker AR, Timmis JN, Orford SJ (2005) Two WD-repeat genes from cotton are functional homologues of the *Arabidopsis thaliana* *TRANSPARENT TESTA GLABRA1 (TTG1)* gene. *Plant Mol Biol* 57:67–81
- Jenkins JN, McCarty JC, Wu J, Saha S, Gutierrez O, Hayes R, Stelly DM (2007) Genetic effect of thirteen *Gossypium barbadense* L. chromosome substitution line in topcrosses with Upland cotton cultivars: II. Fiber quality traits. *Crop Sci* 47:561–570
- Jiang CX, Wright RJ, El-Zik KM, Paterson AH (1998) Polyploid formation created unique avenues for response to selection in *Gossypium* (cotton). *Proc Natl Acad Sci USA* 95:449–4424
- Kamiya T, Kawabe A, Miyashita NT (2002) Nucleotide polymorphism at *Atmyb2* locus of the wild plant *Arabidopsis thaliana*. *Genet Res Camb* 80:89–98
- Kohel RJ, Yu J, Park YH, Lazo GR (2001) Molecular mapping and characterization of traits controlling fiber quality in cotton. *Euphytica* 121:163–172
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen* 12:172–175
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Lacape JM, Nguyen TB, Courtois B, Belot JL, Giband M, Gourelot JP, Gawryziak G, Roques S, Hau B (2005) QTL analysis of cotton fiber quality using multiple *Gossypium hirsutum* × *Gossypium barbadense* backcross generations. *Crop Sci* 45:123–140
- Lee JJ, Hassan OSS, Gao W, Wei NE, Kohel RJ, Chen XY, Payton P, Sze SH, Stelly DM, Chen ZJ (2006) Developmental and gene expression analysis of a cotton naked seed mutant. *Planta* 223:418–432
- Lin Z, He D, Zhang X, Nie Y, Guo X, Feng C, Stewart JMcD (2005) Linkage map construction and mapping QTL for cotton fibre quality using SRAP, SSR and RAPD. *Plant Breed* 124:180–187
- Liu S, Saha S, Stelly DM, Burr B, Cantrell RG (2000) Chromosomal assignment of microsatellite loci in cotton. *J Hered* 91:326–332
- Liu Q, Brubaker CL, Green AG, Marshall DR, Sharp PJ, Singh SP (2001) Evolution of the *FAD-2* fatty acid desaturase 5' UTR intron and the molecular systematics of *Gossypium* (Malvaceae). *Am J Bot* 88:92–102

- Loguercio LL, Zhang JQ, Wilkins TA (1999) Differential regulation of six novel MYB-domain genes defines two distinct expression patterns in allotetraploid cotton (*Gossypium hirsutum* L.) Mol Genet 261:660–671
- Mei M, Syed NH, Gao W, Thaxton PM, Smith CW, Stelly DM, Chen ZJ (2004) Genetic mapping and QTL analysis of fiber-related traits in cotton (*Gossypium*). Theor Appl Genet 108:280–291
- Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
- Oppenheimer OG, Hermin PL, Sivakumaran S, Esch J, Marks DM (1991) A myb gene required for leaf trichome differentiation in *Arabidopsis* is expressed in stipules. Cell 67:483–493
- Park YH, Alabady MS, Ulloa M, Sickler B, Wilkins TA, Yu J, Stelly DM, Kohel RJ, El-Shihy OM, Cantrell RG (2005) Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. Mol Genet Genomics 274:428–441
- Paterson AH, Saranga Y, Menz M, Jiang CX, Wright RJ (2003) QTL analysis of genotype \times environment interactions affecting cotton fiber quality. Theor Appl Genet 106:384–396
- Perez-Rodriguez M, Jaffe FW, Butelli E, Glover BJ, Martin C (2005) Development of three different cell types is associated with the activity of a specific MYB transcription factor in the ventral petal of *Antirrhinum majus* flowers. Development 132:359–370
- Pflieger S, Lefebvre V, Caranta C, Blattes A, Goffinet B, Palloix A (1999) Disease resistance gene analogs as candidates for QTLs involved in pepper-pathogen interactions. Genome 42:1100–1110
- Pflieger S, Lefebvre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7:275–291
- Reinisch AJ, Dong J, Brubaker CL, Stelly DM, Wendel JF, Paterson AH (1994) A detailed RFLP map of cotton, *Gossypium hirsutum* \times *Gossypium barbadense*: chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847
- Ren LH, Guo WZ, Zhang TZ (2002) Identification of quantitative trait loci (QTLs) affecting yield and fiber properties in chromosome 16 in cotton using substitution line. Acta Bot Sin 44:815–820
- Riechmann JL, Heard J, Martin G, Reuber L, Jiang CZ, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu GL (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110
- Rong J, Abbey C, Bowers JE, Brubaker CL, Chang C, Chee PW, Delmonte TA, Ding X, Garza JJ, Marler BS, Park C, Pierce GJ, Rainey KM, Rastogi VK, Schulze SR, Trolinder NL, Wendel JF, Wilkins TA, Williams-Coplin TD, Wing RA, Wright RJ, Zhao X, Zhu L, Paterson AH (2004) A 3347-locus genetic recombinant map of sequence-tagged sites reveals features of genome organization, transmission and evolution of cotton (*Gossypium*). Genetics 166:389–417
- Rong J, Feltus FA, Waghmare VN, Pierce GJ, Chee PW, Draye X, Saranga Y, Wright RJ, Wilkins TA, May OL, Wayne S, Gannaway JR, Wendel JF, Paterson AH (2007) Meta-analysis of polyploid cotton QTLs shows unequal contributions of subgenomes to a complex network of genes and gene clusters implicated in lint fiber development. Genetics 176:2577–2588
- Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19:2496–2497
- Saha S, Jenkins JN, Wu J, McCarty JC, Gutierrez OA, Percy RG, Cantrell RG, Stelly DM (2006a) Effect of chromosome-specific introgression in Upland cotton on fiber and agronomic traits. Genetics 172:1927–1938
- Saha S, Raska DA, Stelly DM (2006b) Upland cotton (*Gossypium hirsutum* L.) \times Hawaiian cotton (*G. tomentosum* Nutt. ex Seem.) F₁ hybrid hypoaneuploid chromosome substitution series. J Cot Sci 12:263–272
- Saranga Y, Menz M, Jiang CX, Wright RJ, Yakir D, Paterson AH (2001) Genomic dissection of genotype \times environment interactions conferring adaptation of cotton to arid conditions. Genome Res 11:1988–1995
- Schiefelbein J (2003) Cell-fate specification in the epidermis: a common patterning mechanism in the root and shoot. Curr Opin Plant Biol 6:74–78
- Schneider K, Schafer-Pregl R, Borchardt DC, Salamini F (2002) Mapping of QTLs for sucrose content, yield and quality in a sugar beet population fingerprinted by EST-related markers. Theor Appl Genet 104:1107–1113
- Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Mol Biol Evol 20:633–643
- Shen X, Guo W, Zhu X, Yuan Y, Yu JZ, Kohel RJ, Zhang T (2005) Molecular mapping of QTLs for fiber qualities in three diverse lines in Upland cotton using SSR markers. Mol Breed 15:169–181
- Shen X, Zhang T, Guo W, Zhu X, Zhang X (2006) Mapping fiber and yield QTLs with main, epistatic, and QTL \times environment interaction effects in recombinant inbred lines of Upland cotton. Crop Sci 46:61–66
- Shen X, Guo W, Lu Q, Zhu X, Yuan Y, Zhang T (2007) Genetic mapping of quantitative trait loci for fiber quality and yield trait by RIL approach in Upland cotton. Euphytica 155:371–380
- Small RL, Wendel JF (2002) Differential evolutionary dynamics of duplicated paralogous Adh loci in allotetraploid cotton (*Gossypium*). Mol Biol Evol 19:597–607
- Small RL, Ryburn JA, Cronn RC, Seelanan T, Wendel JF (1998) The tortoise and the hare: choosing between noncoding plastome and nuclear Adh sequences for phylogeny reconstruction in recently diverged plant group. Am J Bot 85:1301–1315
- Small RL, Ryburn JA, Wendel JF (1999) Low levels of nucleotide diversity at homoeologous Adh loci in allotetraploid cotton (*Gossypium* L.). Mol Biol Evol 16:491–501
- Somers DJ, Kirkpatrick R, Moniwa M, Walsh A (2003) Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 49:431–437
- Stelly DM, Saha S, Raska DA, Jenkins JN, McCarty JC, Gutierrez OA (2005) Registration of 17 Upland (*Gossypium hirsutum*) cotton germplasm lines disomic for different *G. barbadense* chromosome or arm substitutions. Crop Sci 45:2663–2665
- Suo J, Liang X, Pu L, Zhang Y, Xue Y (2003) Identification of *GhMYB109* encoding a R2R3 MYB transcription factor that expressed specifically in fiber initials and elongating fibers of cotton (*Gossypium hirsutum* L.). BBA-Gene Struct Expr 1630:25–34
- Szymanski DB, Lloyd AM, Marks MD (2000) Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. Trends Plant Sci 5:214–219
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustalx windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882
- Van Ooijen JW (2006) JoinMap[®] 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen
- Wang Z, Taramino G, Yang D, Liu G, Tingey SV, Miao GH, Wang GL (2001) Rice ESTs with disease-resistance gene- or defense-response gene-like sequences mapped to regions containing major resistance genes or QTLs. Mol Genet Genomics 265:302–310
- Wang S, Wang JW, Yu N, Li CH, Luo B, Gou JY, Wang LJ, Chen XY (2004) Control of plant trichome development by a cotton fiber MYB gene. Plant Cell 16:2323–2334

- Wang B, Guo W, Zhu X, Wu Y, Huang N, Zhang T (2006a) QTL mapping of fiber quality in an elite hybrid derived-RIL population of upland cotton. *Euphytica* 152:367–378
- Wang K, Song X, Han Z, Guo W, Yu JZ, Sun J, Pan J, Kohel RJ, Zhang T (2006b) Complete assignment of the chromosomes of *Gossypium hirsutum* L. by translocation and fluorescence in situ hybridization mapping. *Theor Appl Genet* 113:73–80
- Wendel JF, Cronn RC (2003) Polyploid and the evolutionary history of cotton. *Adv Agron* 78:139–186
- Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16:2719–2733
- Wu Y, Machado AC, White RG, Llewellyn DJ, Dennis ES (2006) Expression profiling identifies genes expression early during lint fibre initiation in cotton. *Plant Cell Physiol* 47:107–127
- Yang SS, Cheung F, Lee JJ, Ha M, Wei NE, Sze SH, Stelly DM, Thaxton P, Triplett B, Town CD, Chen ZJ (2006) Accumulation of genome-specific transcripts, transcription factors and phytohormonal regulators during early stages of fiber cell development in allotetraploid cotton. *Plant J* 47:761–775
- Zhang J, Guo W, Zhang T (2002) Molecular linkage mapping of allotetraploid cotton (*Gossypium hirsutum* L. × *Gossypium barbadense* L.) with a haploid population. *Theor Appl Genet* 105:1166–1174
- Zhang W, Gianibelli MC, Ma W, Rampling L, Gale KR (2003) Identification of SNPs and development of allele-specific PCR markers for γ -gliadin alleles in *Triticum aestivum*. *Theor Appl Genet* 107:130–138
- Zhang ZS, Xiao YH, Luo M, Li XB, Luo XY, Hou L, Li DM, Pei Y (2005) Construction of a genetic linkage map and QTL analysis of fiber-related traits in upland cotton (*Gossypium hirsutum* L.) *Euphytica* 144:91–99
- Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH (1998) Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res* 8:479–492
- Zhu YL, Song QJ, Hyten DL, Tassell CPV, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. *Genetics* 163:1123–1134