Journal of Heredity 2006:97(5):444–450
doi:10.1093/jhered/esl023
Advance Access publication September 20, 2006

Infraspecific DNA Methylation Polymorphism in Cotton (*Gossypium hirsutum* L.)

ANNA L. KEYTE, RYAN PERCIFIELD, BAO LIU, AND JONATHAN F. WENDEL

From the Department of Biology, Duke University, Durham, NC 27710 (Keyte); the Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011 (Percifield and Wendel); and the Laboratory of Molecular Epigenetics, Northeast Normal University, Changchun 130024, China (Liu).

Address correspondence to J. F. Wendel at the address above, or e-mail: jfw@iastate.edu.

Abstract

Cytosine methylation is important in the epigenetic regulation of gene expression and development in plants and has been implicated in silencing duplicate genes after polyploid formation in several plant groups. Relatively little information exists, however, on levels and patterns of methylation polymorphism (MP) at homologous loci within species. Here we explored the levels and patterns of methylation-polymorphism diversity at CCGG sites within allotetraploid cotton, *Gossypium hirsutum*, using a methylation-sensitive amplified fragment length polymorphism screen and a selected set of 20 *G. hirsutum* accessions for which we have information on genetic polymorphism levels and relationships. Methylation and MP exist at high levels within *G. hirsutum*: of 150 *Hpa*II/*Msp*I sites surveyed, 48 were methylated at the inner cytosine (32%) and 32 of these were polymorphic (67%). Both these values are higher than comparable measures of genetic diversity using restriction fragment length polymorphisms. The high percentage of methylation-polymorphic sites and potential relationship to gene expression underscore the potential significance of MP within and among populations. We speculate that biased correlation of methylation-polymorphic sites and genes in cotton may be a consequence of polyploidy and the attendant doubling of all genes.

The information content of the genome includes not only its primary sequence but also modifications thereof, including methylation of cytosine residues (Tariq and Paszkowski 2004). DNA methylation plays a role in diverse cellular activities, such as carcinogenesis, gene silencing, genomic imprinting, chromatin remodeling, dosage compensation, DNA replication timing, and disease defense and is known to be meiotically as well as mitotically inherited in plants (Finnegan et al. 1996; Pikaard 1999; Bartee et al. 2001; Richards and Elgin 2002; Kato et al. 2003; Meng et al. 2003; Berger 2004; Chan et al. 2004). Methylation occurs predominantly at CG and CNG symmetric sequences, although cytosine methylation at nonsymmetric sites is also common (Gruenbaum et al. 1981; McClelland 1983; Oakeley and Jost 1996). After DNA replication, specific methyltransferases use strand symmetry to guide methylation of the newly synthesized, hemimethylated DNA. In this fashion, epigenetic information in the form of methylation is transferred to daughter strands (Holliday and Pugh 1975). Viroid RNA and small double-stranded RNAs have been shown to serve as signals to activate de novo DNA methylation in plants, which may be followed by transcriptional silencing of homologous genes (Fojtova et al. 2003; Matzke et al. 2004; Steimer et al. 2004). Loss of methylation through knockdown of a methyltransferase results in ectopic gene expression and abnormal development in plants (Finnegan et al. 1996; Ronemus et al. 1996; Jacobsen et al. 2000). On polyploidization, methylation may result in control of the expression of redundant genes and genome stabilization (Lee and Chen 2001; Adams et al. 2003; Scheid et al. 2003). Utilization of cytosine methylation for epigenetic regulation of gene expression could also be advantageous over classic genetic mutation for adaptation in polyploids (Wendel 2000; Lee and Chen 2001; Liu and Wendel 2003).

Gossypium hirsutum L. is a New World allotetraploid represented by a wide range of morphological and ecological forms that vary from highly domesticated, early cropping, high-yielding annuals with abundant fiber to large tropical plants with coarse, sparse fiber (Brubaker et al. 1999; Applequist et al. 2001). Modern cultivars of this species, known as Upland cultivars, dominate world cotton production and have displaced indigenous cottons in the Old World (Lee 1984). Isozyme and restriction fragment length polymorphism (RFLP) data indicate 2 centers of genetic diversity, one in Mesoamerica and the other in the Caribbean, and a strong geographical component to the variation (Wendel et al. 1992; Brubaker and Wendel 1994). In addition, colonial and cultural

influences are evident in patterns of relationships among cotton cultivars and germplasm accessions, as is a profound genetic bottleneck in Upland cultivars associated with the development of the modern high-yielding forms.

Knowledge of epigenetic diversity in *G. hirsutum* would contribute to our understanding of the diversity in cotton while exploring a relatively new area of epigenetic research: methylation variation within species. Here we assess infraspecific levels of methylation polymorphism (MP) by using a methylation-sensitive modification of the common amplified fragment length polymorphism (AFLP) technique (Messeguer et al. 1991; Reyna-Lopez et al. 1997; Xiong et al. 1999; Ashikawa 2001; Liu et al. 2001; Portis et al. 2003) that has been employed in many studies of plant genomes. Methylation-sensitive isoschizomers *Hpa*II and *Msp*I are used in parallel reactions; each recognizes the sequence CCGG but differs in its sensitivity to DNA methylation at the inner cytosine. Using a selected set of *G. hirsutum* cultivars and accessions for which we have previously obtained information on genetic polymorphism levels and relationships (and 2 accessions of *Gossypium barbadense*, a second cultivated cotton species), we here address MP. Specifically, we ask the following: 1) What percentage of the CCGG sites are methylated as determined by differential digestions? 2) How extensive is MP among these sites within cotton? 3) What kinds of genomic sequences are subject to MP?

Table 1. *Gossypium hirsutum* and *Gossypium barbadense* accessions studied

ID	Accession[a]	Geographic origin
1	TX34	Chiapas/Huehuetenango 'latifolium'
2	TX168	Jutiapa 'latifolium'
3	TX98	Chiquimula 'latifolium'
4	TX116	Santa Rosa 'latifolium'
5	TX493	Yucatan/Baja Verapaz 'latifolium'
6	TX6	Puebla/Oaxaca 'latifolium'
7	TX119	El Salvador 'latifolium'
8	TX303	Pacific Coast 'palmeri'
9	TX2089	Southern Mesoamerican 'palmeri'
10	TX230	Zacapa 'punctatum'
11	TX141	Jutiapa
12	TX461	Oaxaca/Puebla/Veracruz
13	TX1009	Santa Rosa/El Salvador
14	TX44	Chiapas
15	TX656	Yucatan/Petén
16	TX706	Honduras/Nicaragua
17	TX724	Belize
18	TX210	Zacapa/Alta Verapaz
19	Delcot344	Upland cultivar
20	Tamcot CAMD-E	Upland cultivar
21	B444	*G. barbadense*
22	B559	*G. barbadense*

[a] Accession numbers are those used in the National Collection of *Gossypium* Germplasm, College Station, TX.

Materials and Methods

Plant Materials and DNA Extraction

The 20 accessions used in this study (Table 1) were selected from the 19 representative groups of *G. hirsutum* that encompass diversity in the species (2 cultivars representing the modern cultivated group), as determined by an earlier RFLP survey of approximately 200 nuclear loci (Brubaker and Wendel 1994). We also included as outgroups 2 representative samples of *G. barbadense*, a second allopolyploid cotton species that is also commercially important (as "pima cotton," "Egyptian cotton"). Genomic DNA was extracted from young expanding leaves using the Qiagen (Valencia, CA) DNA Extraction kit. Care was taken to collect leaves at the same developmental stage so that any developmental variation in methylation, to the extent that it occurs within leaves, would not confound our ability to determine genotype-specific variation in methylation patterns. For a subset of accessions (9 of 22), DNA was extracted from 2 plants of the same accession and used in the methylation-sensitive AFLP (msAFLP) analysis below. This was done to ensure repeatability of banding patterns and methylation stability within accessions.

Methylation-Sensitive Amplified Fragment Length Polymorphism Assay

Our protocol is a combination of the msAFLP methods used by Reyna-Lopez et al. (1997) and the online protocol of Mark E. Berres (University of Wisconsin–Madison, http://ravel.zoology.wisc.edu/sgaap/AFLP_html/AFLP.htm). No more than 300 ng of genomic DNA was digested with 5 U *Hpa*II (50 000 U/ml) or *Msp*I (100 000 U/ml) and 5 U *Eco*RI (100 000 U/ml) in a total volume of 20 μl of the appropriate buffer at 37 °C for 3 h. Rare and frequent cutter digestions (i.e., 4- and 6-bp recognition sites, respectively) were performed simultaneously. Concentrated supplies of the restriction enzymes were used (New England BioLabs, Ipswich, MA) to limit the glycerol content to less than 5% and avoid spurious digestion. The reaction was stopped by heating the mixture to 80 °C for 10 min. Digestion and adapter ligations were carried out separately to avoid generation of concatemers. Adapters were prepared by mixing equal amounts of oligonucleotides "*Eco*RI–adapterI" and "*Eco*RI–adapterII," each at a concentration of 100 μM (Table 2). The mixtures were heated to 95 °C for 5 min and then allowed to anneal by cooling slowly over the course of 25 min. This process was repeated for the *Hpa*II/*Msp*I adapter combination. T4 DNA ligase (200 U, New England BioLabs), 37.5 pmol of each adapter, and the provided ligation buffer were incubated at 16 °C overnight with the digestion product, bringing the total reaction volume to 30 μl. Ligation product was diluted to 15% in sterile water.

Preamplification was performed using 10.0 μl dilute ligation product, 15 pmol each of *Eco*RI + A and *Hpa*II/*Msp*I + 0 primers, 1× polymerase chain reaction (PCR) buffer, 3.0 μl 25 mM $MgCl_2$, 4.0 μl 2.5 mM deoxynucleoside triphosphates (dNTPs), and 2.5 U *Taq* DNA polymerase (GIBCO BRL, Gaithersburg, MD) in a total volume of 50.0 μl. The reaction entailed a 2-min hold at 75 °C, followed by 20 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 2 min, and finally, 10 min at 60 °C before a 10 °C hold. The product was diluted to 5% in sterile water.

Table 2. Primers and adapters used in this study (overhanging nucleotides shown in italics)

Adapters	
*Eco*RI–adapterI	5′-*CTCGTA*GACTGCGTACC
*Eco*RI–adapterII	5′-*AATT*GGTACGCAGTC
*Hpa*II/*Msp*I–adapterI	5′-*GA*TCATGAGTCCTGCT
*Hpa*II/*Msp*I–adapterII	5′-*CG*AGCAGGACTCATGA
Preselective primers	
*Eco*RI + A	5′-GACTGCGTACCAATTCA
*Hpa*II/*Msp*I + 0	5′-ATCATGAGTCCTGCTCGG
Selective primers	
*Eco*RI + AAC	*Eco*RI + A + AC
*Eco*RI + ACA	*Eco*RI + A + CA
*Eco*RI + ACG	*Eco*RI + A + CG
*Eco*RI + AGC	*Eco*RI + A + GC
*Hpa*II/*Msp*I + TCAC	*Hpa*II/*Msp*I + 0 + TCAC
*Hpa*II/*Msp*I + TCAA	*Hpa*II/*Msp*I + 0 + TCAA

The selective amplification was conducted in a volume of 25 μl using 5.0 μl dilute preamplification product from above, 1× PCR buffer, 1.5 μl 25 mM $MgCl_2$, 3.0 μl 2.5 mM dNTPs, 5.0 pmol each of 2 *Eco*RI primers, 20.0 pmol *Hpa*II/*Msp*I primer, 1.0 μl 10 mg/ml bovine serum albumin, and 1.25 U *Taq* DNA polymerase. One of the *Eco*RI primers in the reaction was prelabeled with tetrachlorinated analogue of 6-carboxyfluorescein (6-FAM) and another with 6-FAM for fluorescent visualization of selective amplification products. The PCR parameters included an initial hold at 94 °C for 2 min; 10 touchdown cycles (94 °C for 30 s, 65 °C for 30 s, and 72 °C for 2 min) during which the annealing temperature was decremented 1 °C each cycle; 25 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 2 min; 60 °C for 10 min; and finally a 10 °C hold.

A master mix comprising 2.4 μl formamide, 0.5 μl blue dextran, and 0.6 μl GeneScan-500 [TAMRA] size standard (PE Biosystems, Foster City, CA) was added to 1.5 μl selective amplification product and heated for 2 min at 95 °C and then snap cooled on ice. A final volume of 1.3 μl was electrophoresed on a 4.5% acrylamide gel on an Applied Biosystems (ABI, Foster City, CA) Prism 377 DNA sequencer for 3.0 h. The program GENESCAN was used to capture the fluorescent bands, and the gel image was scored by eye. To ensure repeatability, at least 2 replicate reactions using the same DNA starting material were performed for the entire msAFLP process from digestion to gel scoring for all accessions and all primer sets, in addition to the biological replicates mentioned above.

Obtaining Sequences of Polymorphic Fragments

To determine the identity of fragments that were judged to be polymorphic with respect to methylation status, selective *Eco*RI primers were end labeled with radioisotope (ATP [^{32}P] end-labeling grade from ICN Radiochemicals, Solon, OH) and used in selective amplification reactions identical to those above. The reaction products were run on denaturing 6% polyacrylamide gels, and polymorphic bands were cut from the gel and rehydrated by boiling in 100 μl water for 5 min. The eluted DNA was used as template for PCR after the cycling conditions of the selective amplification. The reactions were carried out in a volume of 25 μl using 5.0 μl eluted DNA, 1× PCR buffer, 0.75 μl 50 mM $MgCl_2$, 3.0 μl 2.5 mM dNTPs, 10.0 pmol *Eco*RI + A primer, 10.0 pmol *Hpa*II/*Msp*I primer, and 1.0 U *Taq* DNA polymerase. This PCR product was purified using the Qiagen PCR purification kit and then directly sequenced using the *Eco*RI + A primer. Sequences obtained are deposited in GenBank under accession numbers AY789113–AY789119.

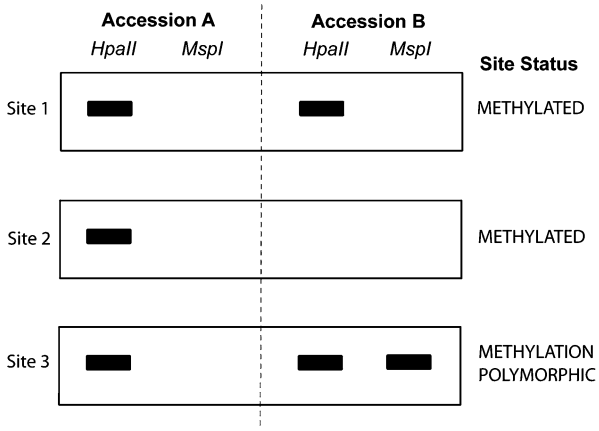

Figure 1. Scheme used to classify sites as methylated or methylation polymorphic among accessions, using as an example 2 arbitrary accessions termed A and B.

Data Analysis

The BLAST package on the National Center for Biotechology Information server was used to search public databases for sequences similar to those we obtained, as well as to a growing collection of cotton expressed sequence tags (SF Altschul, W Gish, W Miller, EW Myers, and DJ Lipman, unpublished data). The statistics package NTSYSpc version 2.10p (Applied Biostatistics, Setauket, NY) was used for principal component analysis (PCA) of the MP data using a correlation matrix of the polymorphic fragments among all accessions. Data were visualized by projection of the accessions onto a plane defined by the first two principal components.

Results

A CCGG site for a particular accession was classified as "methylated" if a band was present in either the *Msp*I or *Hpa*II lane but not the other, "not methylated" if bands were present in both lanes, or "unknown" if bands were absent in both lanes (Figure 1). A site was considered "methylation polymorphic" (MP) if there was at least one accession in which the site was methylated and at least one accession for which the site was not methylated. *Msp*I and *Hpa*II are both sensitive to methylation at the outer cytosine of the CCGG recognition sequence; therefore, absence of bands in both the *Msp*I and *Hpa*II lanes could be due to either genetic polymorphism or hypermethylation.

Sites were scored and included in the analysis only if replicate AFLP reactions showed the same band pattern.

Table 3. Presence or absence of methylation at 34 sites for 20 *Gossypium hirsutum* and 2 *Gossypium barbadense* accessions[a]

	TX number/accession number																						
ID no.	34	168	98	116	493	6	119	303	2089	230	141	461	1009	44	656	706	724	210	Delcott 344	Tamcott CAMD-E	B444	B559
G1.384	1	1	1	1	1	1	?	1	1	1	1	0	1	1	1	0	0	?	1	1	1	1
G1.367	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	?	?
G1.307	0	0	0	0	0	0	0	1	0	0	0	0	0	0	?	0	0	0	0	0	?	?
G1.192[b]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
G2.428	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
G2.151[b]	0	0	0	1	0	?	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
G3.491	1	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	1
G3.482	0	1	1	0	1	?	1	0	0	1	1	1	0	0	0	1	0	0	0	0	?	?
G3.430	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
G3.392[c]	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
G3.292	0	0	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
G3.274[b]	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G3.160[b]	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1
G3.154	1	1	1	1	1	?	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
G4.491[b]	1	0	1	1	1	?	1	?	1	1	1	1	1	1	0	1	1	1	1	1	1	1
G4.368	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1
G4.288	?	1	1	1	1	1	1	?	0	1	?	0	?	1	?	0	1	1	1	1	?	?
G4.194	1	?	0	0	?	1	?	1	1	0	1	1	0	?	1	1	?	?	?	?	?	0
G4.172	0	0	0	0	1	0	1	0	?	0	0	0	0	0	1	0	0	0	0	0	?	?
G4.165	0	1	1	1	1	1	1	1	1	1	0	0	0	0	?	?	?	1	0	0	0	0
B1.424	1	?	1	1	?	?	?	0	?	1	0	0	0	1	1	1	1	1	1	1	0	?
B1.218	0	0	0	1	0	?	?	0	0	0	0	0	0	0	0	0	0	?	0	1	?	?
B2.493	0	0	0	0	0	1	1	0	0	0	0	0	0	?	0	0	0	?	0	0	0	?
B2.465	?	?	?	?	?	1	?	?	?	?	0	?	?	?	?	?	?	?	?	?	?	?
B2.232[c]	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
B2.158	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B3.416	0	1	1	0	?	?	0	?	1	0	0	1	?	1	0	?	0	?	0	?	0	?
B3.359[b]	0	1	0	0	1	0	1	1	1	1	0	1	0	1	1	0	0	0	1	1	1	1
B3.208[b]	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0
B4.499	0	0	0	0	0	1	1	?	0	0	0	0	0	0	1	0	0	?	0	0	0	1
B4.327	?	0	0	?	?	?	?	?	?	?	0	?	1	0	?	?	?	?	?	?	0	0
B4.325	0	?	?	0	0	1	1	1	0	0	0	0	?	?	0	0	0	0	0	0	?	?
B4.175	1	?	1	1	0	?	0	0	0	0	?	0	0	?	0	1	0	?	1	1	1	0

[a] A "1" indicates methylation; "0" indicates no methylation; and "?" indicates that the methylation state is unknown.

[b] Sites from which sequence data were obtained.

[c] Sites that are only methylation polymorphic when *G. barbadense* is included.

The biological replicates (msAFLP analysis on DNA from 2 different individuals of the same accession) demonstrated that there is little diversity within any given accession, and the few differences found were all genetic, not epigenetic.

The primer combinations used yielded 150 sites that could be scored with high confidence and repeatability. Of these 150, 48 were methylated (32%), of which 32 were polymorphic (67%) (Table 3). These numbers do not include 2 sites that were methylated or MP only within the 2 *G. barbadense* accessions examined.

To explore whether patterns of MP mirrored genetic relationships as determined previously (Brubaker and Wendel 1994), a neighbor-joining tree was constructed and visually compared with the earlier tree based on RFLP data. No evident relationships among the 22 accessions studied here were revealed (not shown). However, a PCA performed on the data in Table 3 reveals possible relationships with geography and *G. hirsutum* "race" (Figure 2). The 2 Upland cottons emerged close to one another in multivariate space, and the representative accessions from Belize, Honduras, and El Salvador clustered together. Accessions were coded according to the geographical groups used by Wendel et al. (1992), and those from the same region were often close in multivariate space (see Figure 2A). Seven of the 22 accessions were of race 'latifolium'; these form a cloud that excludes most other accessions (Figure 2B). Members of race 'palmeri' and race 'marie galante' were also adjacent relative to other accessions. The 2 *G. barbadense* accessions did not show significant dissimilarity from the *G. hirsutum* accessions with regard to methylation of the sites considered here. There were significant genetic differences, and because a loss of bands in both the *Hpa*II and *Msp*I lanes was scored as an unknown methylation status, this most likely contributed to an apparent decrease in MP between the 2 species.

To assess the types of genomic sequences subject to MP, we gel-isolated selected AFLP bands and sequenced them.

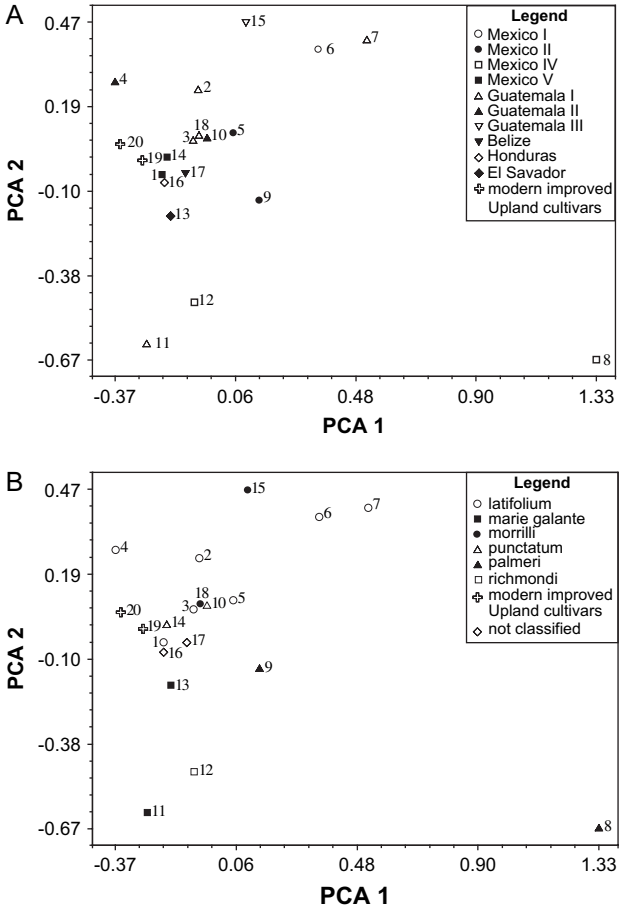

Figure 2. Principal component analysis. (**A**) Data coded by geographical region as in Wendel et al. (1992). (**B**) Coded for racial designation. The first 2 principal components account for 18.7% and 14.6% of the total variance. Identifying numbers next to each data point correspond to accessions as shown in Table 1.

We were able to obtain sequences for 7 polymorphic bands (Table 4), and of these 7, 4 were similar to existing sequences in public databases. These 4 sequences were all similar to plant protein-coding genes or cDNAs, as specified in Table 4.

Discussion

Cytosine methylation has proved to be an important factor in the epigenetic regulation of gene expression, and as such, a number of studies have investigated the role of methylation in silencing duplicate genes after polyploid formation (Comai et al. 2000; Liu et al. 2001; Shaked et al. 2001; Madlung et al. 2002; Salmon et al. 2005). Here we explored the diversity of cytosine methylation at CCGG sites within allotetraploid cotton, *G. hirsutum*. Our results show that methylation as well as MP is widespread within *G. hirsutum*. Of 150 *Hpa*II/*Msp*I sites, 48 were methylated at the inner cytosine (32%) and 32 of these were polymorphic (67%), despite the relatively small number of accessions ($n = 20$) examined. Although there are no comparable studies from other plants, preliminary surveys in *Arabidopsis* (Cervera et al. 2002; Riddle and Richards 2002), rice (Ashikawa 2001; Wang et al. 2004), and *Pisum* (Knox and Ellis 2001) suggest that our results will not be unique, but instead that MP will turn out to be widespread within plant species.

It is noteworthy that previous surveys (Wendel et al. 1992; Brubaker and Wendel 1994) have demonstrated that *G. hirsutum* contains relatively low levels of genetic variation, yet in spite of this a great deal of MP was observed. Brubaker and Wendel (1994) found that 22.4% of RFLP bands are polymorphic, yet 67% of methylated sites are polymorphic, underscoring the potential significance of MP within and among populations. This, together with stability of methylation patterns within a given cotton accession (biological replicates), makes it possible that the MPs may serve as epigenetic markers for certain populations and cultivars. Additional experimentation is needed to confirm the extent and scope of MP, as well as the mode of inheritance of the accession-specific methylation patterns within a broader sampling of this and other species.

The PCA of the MP data reveals some noteworthy similarities to the results of the study of Brubaker and Wendel (1994) of genetic diversity within *G. hirsutum*. They found that whereas the Upland accessions formed a distinct group, accessions from Mexico and Guatemala failed to cluster and were interspersed with one another. The same patterns were found using the MP data (Figure 2A). Brubaker and Wendel also found that 'latifolium' and 'palmeri' races from

Table 4. BLAST search results for sequenced methylation-polymorphic fragments. The size of the sequenced fragment was dictated by the relative location of *Eco*RI and *Hpa*II/*Msp*I cut sites

Band ID	Primers	Size (bp)	Sequence similarity	Length of sequence similarity	Mismatches
G1.192	HM+TCAA/Eco+ACA	192	*Arabidopsis thaliana* serine/threonine protein kinase putative mRNA	74	8
G2.151	HM+TCAA/Eco+AGC	151	*Gossypium arboreum* EST, *A. thaliana* mRNA for 3-phosphoserine phosphatase	76, 75	5, 18
G3.274	HM+TCAC/Eco+ACA	274	*Gossypium hirsutum* cDNA, *A. thaliana* phosphofructokinase family mRNA	91, 85	5, 19
G3.160	HM+TCAC/Eco+ACA	160	*Arabidopsis thaliana* putative aldehyde dehydrogenase	40	7
G4.491	HM+TCAC/Eco+AGC	491	None		
B3.359	HM+TCAC/Eco+AAC	359	None		
B3.208	HM+TCAC/Eco+AAC	208	None		

diverse locations clustered and were associated with the Upland cottons. The PCA in Figure 2B shows a similar relationship. These results suggest that, in general, methylation diversity mirrors genetic diversity; that accession-specific methylation patterns are heritable; and that the pace of MP evolution is not so rapid that recent history of epigenetic evolution is obscured. In *Arabidopsis* (Cervera et al. 2002) and rice (Ashikawa 2001), MP was found not to be related to genetic relatedness among the accessions studied.

Additional clues into the potential importance of MP were provided by obtaining sequence for a subset of the MP bands. For our data set, MP is distinctly correlated with protein-coding genes. Whereas the number of sequences retrieved was small relative to the number of MP sites, it is noteworthy that 4 of the 7 sequences generated had sequence matches to putative genes, despite the fact that the AFLP technique and the enzymes employed do not enrich for the genic component of the genome. Moreover, it was demonstrated recently that, in contrast to mammals, most plant coding sequences are usually not methylated (Rabinowicz et al. 2003). It thus is tempting to speculate that this potential biased correlation of MP sites and genes in cotton is likely a direct consequence of allopolyploidy and the attendant doubling of all genes. In recent years it has become clear that genome doubling leads to extensive alteration in gene expression (Lee and Chen 2001; Adams et al. 2003; Liu and Wendel 2003), including gene silencing, some of which may be mediated through methylation. A natural extension of our study would be to examine the effects of methylation on expression of the genes identified here, as well as to examine the extent of methylation and MP in the parental diploids.

Because some gene-silencing events arise immediately with the onset of polyploid formation, the methylation-polymorphic sites revealed in this study raise questions about the pace of MP accumulation on an evolutionary time scale (Adams et al. 2003; Liu and Wendel 2003; Adams and Wendel 2005). After an initial speciation event, how do accession-specific methylation patterns originate and at what pace do MP sites evolve, as a species diversifies and spreads over a landscape? Answers to this and the correlated question of reversibility are likely to emerge from expanded surveys in this and other species, particularly if combined with independently generated genetic or phylogeographic data. Although the PCA results presented here suggest that patterns of methylation diversity correlate with patterns of genetic diversity, mapping methylation changes onto genetically based phylogeographic trees could diagnose the level within a phylogenetic hierarchy at which any given polymorphism originated; whether a particular polymorphism has a single or multiple origin; whether methylations are subject to high rate of reversion; and by extension of both the pace of MP evolution and its evolutionary liability.

Acknowledgments

Funding for this research was provided by a grant from the USDA-NRI (to J.F.W.). The authors wish to thank Keith Adams and Olivier Fedrigo for helpful discussion, as well as 2 anonymous reviewers.

References

Adams KL, Cronn R, Percifield R, Wendel J. 2003. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100:4649–4654.

Adams KL, Wendel JF. 2005. Novel patterns of gene expression in polyploid plants. Trends Genet 21:539–543.

Applequist WL, Cronn R, Wendel JF. 2001. Comparative development of fiber in wild and cultivated cotton. Evol Dev 3:3–17.

Ashikawa I. 2001. Surveying CpG methylation at 5′-CCGG in the genomes of rice cultivars. Plant Mol Biol 45:31–39.

Bartee L, Malagnac F, Bender J. 2001. Arabidopsis *cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. Genes Dev 15:1753–1758.

Berger F. 2004. Imprinting—a green variation. Science 303:483–485.

Brubaker CL, Bourland FM, Wendel JF. 1999. The origin and domestication of cotton. In: Smith WC, editor. Cotton: origin, history, technology, and production. New York: John Wiley & Sons. p 3–31.

Brubaker CL, Wendel JF. 1994. Reevaluating the origin of domesticated cotton (*Gossypium hirsutum*; Malvaceae) using nuclear restriction fragment length polymorphisms (RFLPs). Am J Bot 81:1309–1326.

Cervera MT, Ruiz-Garcia L, Martinez-Zapater JM. 2002. Analysis of DNA methylation in *Arabidopsis thaliana* based on methylation-sensitive AFLP markers. Mol Genet Genomics 268:543–552.

Chan SW-L, Zilberman D, Xie Z, Johansen LK, Carrington JC, Jacobsen SE. 2004. RNA silencing genes control *de novo* DNA methylation. Science 303:1336.

Comai L, Tyagi AP, Holmes-Davis KWR, Reynolds SH, Stevens Y, Byers B. 2000. Phenotypic instability and rapid genome silencing in newly formed arabidopsis allotetraploids. Plant Cell 12:1551–1567.

Finnegan EJ, Peacock WJ, Dennis ES. 1996. Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. Proc Natl Acad Sci USA 93:8449–8454.

Fojtova M, Houdt HV, Depicker A, Kovarik A. 2003. Epigenetic switch from posttranscriptional to transcriptional silencing is correlated with promoter hypermethylation. Plant Physiol 133:1240–1250.

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A. 1981. Sequence specificity of methylation in higher plant DNA. Nature 292:860–862.

Holliday R, Pugh JE. 1975. DNA modification mechanisms and gene activity during development. Science 187:226–232.

Jacobsen SE, Sakai H, Finnegan EJ, Cao X, Meyerowitz EM. 2000. Ectopic hypermethylation of flower specific genes in *Arabidopsis*. Curr Biol 10:179–186.

Kato M, Miura A, Bender J, Jacobsen SE, Kakutani T. 2003. Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. Curr Biol 13:421–426.

Knox MR, Ellis THN. 2001. Stability and inheritance of methylation states at *Psi*I sites in *Pisum*. Mol Genet Genomics 265:497–507.

Lee H-S, Chen ZJ. 2001. Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids. Proc Natl Acad Sci USA 98:6753–6758.

Lee JA. 1984. Cotton as a world crop. In: Kohel RJ, Lewis CL, editors. Cotton, Agronomy monograph 24. Madison, WI: Crop Science Society of America. p 1–25.

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF. 2001. Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome 44:321–330.

Liu B, Wendel JF. 2003. Epigenetic phenomena and the evolution of plant allopolyploids. Mol Phylogenet Evol 29:365–379.

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L. 2002. Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic *Arabidopsis* allotetraploids. Plant Physiol 129:733–746.

Matzke MA, Aufsatz W, Kanno T, Daxinger L, Papp I, Mette MF, Matzke AJ. 2004. Genetic analysis of RNA-mediated transcriptional gene silencing. Biochim Biophys Acta 1677:129–141.

McClelland M. 1983. The frequency and distribution of methylatable DNA sequences in leguminous plant protein coding genes. J Mol Evol 19: 346–354.

Meng L, Bregitzer P, Zhang S, Lemaux PG. 2003. Methylation of the exon/intron region in the *ubi1* promoter complex correlates with transgene silencing in barley. Plant Mol Biol 53:327–340.

Messeguer R, Ganal MW, Stevens JC, Tanksley SD. 1991. Characterization of the level, target sites and inheritance of cytosine methylation in tomato nuclear DNA. Plant Mol Biol 16:753–770.

Oakeley EJ, Jost J-P. 1996. Nonsymmetrical cytosine methylation in tobacco pollen DNA. Plant Mol Biol 31:927–930.

Pikaard CS. 1999. Nucleolar dominance and silencing of transcription. Trends Plant Sci 4:478–483.

Portis E, Acquadro A, Comino C, Lanteri S. 2003. Analysis of DNA methylation during germination of pepper (*Capsicum annuum* L.) seeds using methylation-sensitive amplification polymorphism (MSAP). Plant Sci 166: 169–178.

Rabinowicz PD, Palmer LE, May BP, Hemann MT, Lowe SW, McCombie WR, Martienssen RA. 2003. Genes and transposons are differentially methylated in plants, but not in mammals. Genome Res 13:2658–2664.

Reyna-Lopez G, Simpson J, Ruiz-Herrera J. 1997. Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. Mol Gen Genet 253: 703–710.

Richards EJ, Elgin SC. 2002. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell 108:489–500.

Riddle NC, Richards EJ. 2002. The control of natural variation in cytosine methylation in *Arabidopsis*. Genetics 162:355–363.

Ronemus MJ, Galbiati M, Ticknor C, Chen J, Dellaporta SL. 1996. Demethylation-induced developmental pleiotropy in *Arabidopsis*. Science 273:654–657.

Salmon A, Ainouche ML, Wendel JF. 2005. Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae). Mol Ecol 14:1163–1175.

Scheid OM, Afsar K, Paszkowski J. 2003. Formation of stable epialleles and their paramutation-like interaction in tetraploid *Arabidopsis thaliana*. Nat Genet 34:450–454.

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13:1749–1759.

Steimer A, Schob H, Grossniklaus U. 2004. Epigenetic control of plant development: new layers of complexity. Curr Opin Pant Biol 7:11–19.

Tariq M, Paszkowski J. 2004. DNA and histone methylation in plants. Trends Genet 20:244–251.

Wang YM, Lin XY, Dong B, Wang YD, Liu B. 2004. DNA methylation polymorphism in a set of elite rice cultivars and its possible contribution to inter-cultivar differential gene expression. Cell Mol Biol Lett 9:543–556.

Wendel JF. 2000. Genome evolution in polyploids. Plant Mol Biol 42: 225–249.

Wendel JF, Brubaker CL, Percival AE. 1992. Genetic diversity in *Gossypium hirsutum* and the origin of upland cotton. Am J Bot 79:1291–1310.

Xiong LZ, Xu CG, Maroof MAS, Zhang Q. 1999. Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. Mol Gen Genet 261:439–446.

Received April 22, 2005
Accepted July 31, 2006

Corresponding Editor: James Hamrick