Coordinated and Fine-Scale Control of Homoeologous Gene Expression in Allotetraploid Cotton

LEX E. FLAGEL, LIPING CHEN, BHUPENDRA CHAUDHARY, AND JONATHAN F. WENDEL

From the Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA 50011.

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

Within polyploid plant species, it has been demonstrated that homoeologous genes (genes duplicated by polyploidy) often display dynamic expression patterns. To determine if chromosomal location plays a role in establishing these expression patterns, we analyzed the relative levels of homoeolog expression among linked genes from 2 locations in the cotton genome. Genes from the region containing the alcohol dehydrogenase A gene show coordinated expression across several tissues, whereas genes from the region containing cellulose synthase A do not. These results indicate that changes in homoeolog expression may be constrained by linkage in some genomic regions, whereas in other regions, homoeolog expression is largely decoupled from physical proximity. Furthermore, these results suggest that both large- and small-scale regulatory mechanisms may control homoeolog expression patterns.

Key words: cotton, duplicate gene expression, Gossypium hirsutum, linkage, polyploidy

Polyploidy is a common and phylogenetically widespread phenomenon in angiosperms. Many polyploids are chromosomal “allopolyploids,” typically descendants of a polyploidy event that included interspecific hybridization (Wendel and Doyle 2005). As such, allopolyploids maintain 2 or more divergent genomes within a common nucleus. This sharing of a common nucleus creates the potential for a variety of intergenomic interactions, including homoeologous recombination (Udall et al. 2005; Gaeta et al. 2007), genomic deletions (Ozkan et al. 2001; Kashkush et al. 2002), and modifications of homoeologous gene expression (Adams et al. 2003; Mochida et al. 2003; Bottley et al. 2006; Flagel et al. 2008; Hovav et al. 2008). These findings have led to the speculation that interactions among coresident genomes play a key role in the development of novel phenotypes in allopolyploid species (Wendel 2000; Osborn et al. 2003; Chen 2007).

One of the more impressive discoveries regarding allopolyploids concerns the scale and scope of unequal contributions to the transcriptome made by homoeologous genes. Adams et al. (2003) showed that among 40 homoeologous gene pairs in cotton, 10 exhibit biased expression patterns, including developmentally regulated expression ratio variation among homoeologs and reciprocal silencing of alternative homoeologs in different floral whorls. Similar findings were reported in wheat (Mochida et al. 2003; Bottley et al. 2006). More recently, Flagel et al. (2008) and Hovav et al. (2008) used microarray technology to show that homoeolog expression biases are common throughout the transcriptome and can vary temporally and developmentally, even within the single-celled trichomes of Gossypium (cotton “fibers”).

Despite these advances into understanding the phenomenon of homoeolog expression variation, the responsible molecular mechanisms remain largely unexplored and unknown. Although both cis and trans factors are implicated (Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF, unpublished data), little is understood regarding the roles of genetic and epigenetic factors in creating homoeolog expression biases, nor how these expression changes affect adjacent genes. With respect to the latter, preliminary evidence from allotetraploid Arabidopsis indicates that homoeolog expression may be quite variable even among closely spaced genes on a single bacterial artificial chromosome (BAC) (Lee and Chen 2001). To explore this suggestion further and in a different system, we quantitatively assayed homoeolog expression ratios for genes found on 2 well-characterized regions of the allotetraploid cotton (Gossypium hirsutum) genome. By collecting homoeolog expression data from a wide range of tissues, we sought to determine the effect of physical proximity on homoeolog expression. If closely linked genes show correlated homoeolog expression biases, one might infer that the biases likely result from the action of a single cis-acting factor such as a chromatin modification or a shared enhancer affecting a large genomic region. Conversely, if linked genes show uncorrelated expression patterns, one would conclude that multiple fine-scale factors may be at work, including localized divergence among homoeologous cis-regulatory regions or epigenetic marks.
**Materials and Methods**

*Gossypium* has emerged as an excellent system for studying patterns of homoeologous gene expression. This genus contains 5 extant allotetraploid species that were formed during a single polyploidization event approximately 1–2 million years ago (Senchina et al. 2003). Allotetraploids (AD genome) combine diploid genomes from the “A” and “D” species groups, are chromosomally diploidized (i.e., form bivalents at meiosis), and have a genome size that is approximately additive with respect to those of its diploid progenitors (Wendel and Cronn 2003). Beyond their well-understood origin and phylogeny, there are also ample genetic and genomic resources for cotton. Of importance to this study are annotated BAC sequences from *G. hirsutum* for both homoeologous genomes around the regions containing the alcohol dehydrogenase A (*AdhA*; GenBank accessions EF457753 and EF457754) and the cellulose synthase A genes (*CesA*; GenBank accessions AY632359 and AY632360) (Grover et al. 2004, 2007). By aligning genes on these BACs, we were able to identify single nucleotide polymorphism (SNP) markers specific to the homoeologous A and D genomes for 3 genes from the *AdhA* BACs and 8 genes from the *CesA* BACs (Figure 1A). These genome-specific SNPs were used as targets for the MassARRAY (Sequenom, San Diego, CA) mass spectrometry platform, which can quantitatively assay relative homoeolog transcript abundance as described below. Our goal was to ascertain the relative ratios of homoeolog expression in a diverse panel of tissues and in particular to evaluate the degree to which gene expression among homoeologs is correlated with genomic proximity.

Following protocols described elsewhere (Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF, unpublished data), we extracted RNAs and prepared cDNA libraries from 18 tissue types (Figure 1B). These cDNA libraries were then subjected to MassARRAY mass spectrometry, which utilizes differences in mass

---

**Figure 1.** Representations of *AdhA* and *CesA* genomic regions from *Gossypium hirsutum* and summary of relative homoeolog contribution to the transcriptome for neighboring genes in multiple tissues. (A) Arrows represent genes for which expression data were generated from the *AdhA* and *CesA* BACs. The direction of the arrow indicates the direction of transcription for each gene, and the number of bases between genes is given for both the A and D genomes of the allopolyploid. The order and direction of transcription is conserved between the A and D genomes for all genes used in this study. Asterisks above the *CesA* BAC mark intervals that contain a predicted gene which could not be assayed due to lack of genome-specific SNPs. (B) Mean expression results for the *AdhA* and *CesA* BACs for each tissue. For each panel, genes are arrayed in physical order (x axis), and homoeolog expression ratio is shown on the y axis (as proportion of the transcriptome contributed by the D genome). Some low-quality assays were excluded from the *CesA* data. dpa, days post anthesis; FAD-Ox, FAD-dependent oxidoreductase; PDI, protein disulfide isomerase; Perm, permease; PPR, pentatricopeptide repeat protein.
between genome-specific SNP variants to quantify the relative abundance of homoeologous transcripts. The resulting homoeolog expression ratio data were then filtered based on internal MassARRAY technical controls to remove assays flagged as “Bad Spectra” or having a frequency of uncertainty >0.2 or an unused extension primer frequency >0.5. Next, assays beyond this expected 1:1 ratio (±25%) were individually excluded. Finally, only those genome values. Any assays beyond this expected 1:1 ratio genomic DNA should yield approximately 1:1 A to D genome values. Any assays beyond this expected 1:1 ratio (±25%) were individually excluded. Finally, only those assays with 3 or more replicates (from a maximum of 9 replicates) are summarized in Figure 1B.

**Results and Discussion**

Several interesting patterns emerge from the homoeolog expression data. First, for nearly all tissues (except anther, leaf lamina, and calyx), genes along the *AdhA* BAC show a consistent V-shaped pattern, in which *AdhA* and *protein disulfide isomerase* consistently maintain a bias toward greater D genome transcription than does *FAD-dependent oxidoreductase*. The maintenance of this pattern is striking and consistent, despite the appreciable differences in homoeolog ratios among tissues (e.g., root vs. 10 days post anthesis fiber). These findings from the *AdhA* region indicate that homoeologous gene expression can be coordinated among adjacent genes. This coordination involves the maintenance of a specific pattern of homoeolog expression bias across many tissues, suggesting the operation of a mechanism that is developmentally persistent and widespread with regard to tissue type.

To our knowledge, this is the first time that large-scale (~40 kb) coordinated control of expression for linked homoeologous genes across several tissues has been demonstrated. Given the breadth of tissues examined, it is reasonable to speculate that this pattern is caused by a developmentally stable, coordinated chromatin modification that exercises control over homoeologous gene expression in most tissues. This regional effect likely dominates specific gene effects in the *AdhA* region.

In contrast to coordinated gene expression in the *AdhA*-region, a different story emerges for genes along the *CesA*-BAC. These genes show no clear pattern of coordinated expression (Figure 1B). Instead, genes in this genomic location suggest fine-scale control of homoeolog expression, including, in several cases, apparently complete gene silencing. For example, in the staminal tube, the D genome contributes nearly all of the transcripts of expressed hypothetical protein 1, whereas its closest neighbor, *G-protein beta*, is predominantly expressed by the A genome. In some cases, genes only a few kilobases apart show very different expression states. For example, in the leaf midrib, *leucine-rich repeat protein 2* expression is approximately equivalent for both genomes, but expressed hypothetical protein 2, which is less than 4 kb away, is represented almost entirely by the A genome homoeolog. Furthermore, we performed statistical analyses on the *CesA*-region to determine if physical distance, gene order, and gene orientation had an effect on pairwise correlations of homoeologous gene expression across tissues. The significance of these physical factors was assessed using permutation tests with physical distance measured in base pairs and tested for both the A and D genome, gene order measured as the number of intervening genes, and gene orientation encoded as either parallel (→→ or ←←) or opposing (←→ or →←). These distance data sets were subjected to 1000 random permutations and compared with the actual correlation. All physical factors were found to be nonsignificantly correlated with expression for the *CesA*-region (physical distance: A genome *P*=0.73; D genome *P*=0.74; gene order *P*=0.74; and gene orientation *P*=0.13). Thus, within this region, our results reveal the presence of localized and intricate regulatory control among homoeologous genomes. This suggests a reduced role for large-scale factors such as heterochromatin-related effects, instead suggesting locally operating factors such as cis-regulatory differences or alteration in localized epigenetic states.

The foregoing explorations of homoeologous gene expression in a diverse panel of tissues reveal 2 distinct patterns, enhancing our understanding of the phenomenon of homoeolog expression bias and providing clues into its genesis. Results from the *AdhA*-region demonstrate that genomic locale may exercise a profound effect on homoeolog expression in a persistent and developmentally widespread manner. This finding is supported by genomic expression analyses in diploids, which often show that angiosperm genomes contain many short tracts of genes (generally less than 10) with correlated expression patterns (Williams and Bowles 2004; Ren et al. 2005, 2007; Zhan et al. 2006; Quesada et al. 2008). Though this phenomenon has not been demonstrated in *Gossypium*, it is possible that expression in the *AdhA*-region for both the A and D genomes falls into one of these coexpressed tracts. Our data add to this body of research by showing evidence of apparent coordination between homoeologous genomes. In contrast, the *CesA*-region demonstrates that closely linked genes can deviate significantly in their homoeologous contribution to the transcriptome. These results imply that widespread patterns of homoeolog expression biases in allotetraploid cotton and wheat (Adams et al. 2003; Mochida et al. 2003; Bottley et al. 2006; Flagel et al. 2008; Hovav et al. 2008) are likely the product of both fine-scale local regulation as well as more far-reaching chromosomal factors.

**Funding**

National Research Initiative of the USDA Cooperative State Research, Education, and Extension Service (2005-35301-15700 to J.F.W.); National Science Foundation (0638418 to J.F.W.).
Acknowledgments

We thank Nathan Springer and Bob Stupar for helpful comments and technical assistance in developing the Sequenom platform for use in cotton. Also we thank the University of Minnesota BioMedical Genomics Center for processing all Sequenom assays.

References


Received December 3, 2008; Revised February 3, 2009; Accepted February 3, 2009

Corresponding Editor: John Burke