Histone modifications define expression bias of homoeologous genomes in allotetraploid cotton

Dewei Zheng, Wenxue Ye, Qingxin Song, Fangpu Han, Tianzhen Zhang, and Z. Jeffrey Chen

1 State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, 1 Weigang Road, Nanjing 210095, China
2 Department of Molecular Biosciences and Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX 78712, USA
3 Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, 1 West Beichen Road, Beijing 100101, China

Author contributions: D.Z., W.Y., and Z.J.C. conceived the research; D.Z. collected most of the data; D.Z., W.Y., and Q.S. performed the data analysis; F.H and T.Z. provided technical assistances and materials; D.Z., W.Y., and Z.J.C. wrote the article.

These authors contributed equally to this work

* Corresponding author: zjchen@austin.utexas.edu
Funding information: The work was supported by grants from the National Natural Science Foundation of China (31290213 and 91631302), National Science Foundation of U.S.A. (IOS1025947 and IOS1444552), Research and Education Innovation Consortium of Jiangsu Province (2013-2015), Jiangsu Collaborative Innovation Center for Modern Crop Production (2014-2016), and The Fundamental Research Funds for the Central Universities (KYRC201204 and KYZ201202-1). The authors declare no competing interest of this work.

Running title: Chromatin and gene expression in tetraploid cotton

Key words: histone modifications, epigenome, transcriptome, polyploidy, evolution, cotton
ABSTRACT

Histone modifications regulate gene expression in eukaryotes, but their roles in gene expression changes in interspecific hybrids or allotetraploids are poorly understood. Histone modifications can be mapped by immunostaining of metaphase chromosomes at the single cell level and/or by chromatin immunoprecipitation-sequencing (ChIP-seq) for individual genes. Here we examined H3K4me3 density and transcriptome maps in root-tip cells of allotetraploid cotton (Gossypium hirsutum L.). The overall H3K4me3 levels were relatively equal between A and D chromosomes, which were consistent with equal numbers of expressed genes between the two subgenomes. However, intensities per chromosomal area were nearly twice as high in the D homoeologs as in the A homoeologs. Consistent with the cytological observation, ChIP-seq analysis showed that more D-homoeologs with biased H3K4me3 levels than A-homoeologs with biased modifications correlated with the greater number of the genes with D-biased expression than that with A-biased expression in most homoeologous chromosome pairs. Two chromosomes displayed different expression levels compared with other chromosomes probably because of translocations, which may affect the local chromatin structure and expression levels for the genes involved. This example of genome-wide histone modifications that determine expression bias of homoeologous genes in allopolyploids provides a molecular basis for the evolution and domestication of polyploid species including important crops.
INTRODUCTION

Chromatin structure is a repeating unit of nucleosomes, each consisting of approximately 147 base pairs of DNA wrapping around a histone octamer with 2 copies each of H2A, H2B, H3, and H4 core histones (Kornberg, 1974; Luger et al., 1997). The core histone tails can be modified to mediate gene expression, growth, and development in plants and animals (Jenuwein and Allis, 2001; Berger, 2007; Li et al., 2007; Liu et al., 2010; Ha et al., 2011). Some modifications, such as histone H3 lysine K9 acetylation (H3K9ac) and lysine 4 trimethylation (H3K4me3), are euchromatic marks and often associated with active transcription, whereas other modifications, such as H3K9me2 and H3K27me3, are known as heterochromatic marks and related to gene repression (Jenuwein and Allis, 2001; Li et al., 2007).

Changes in chromatin structure and consequently gene expression can occur during development and/or in response to internal (“genome shock”) and external (environmental) signals (McClintock, 1984; Chen, 2007; Chen and Tian, 2007). This genome shock is often associated with interspecific hybridization and polyploidization (McClintock, 1984; Comai, 2005; Chen, 2007). The additional genetic materials in new polyploid species can facilitate adaptation to local environments as well as for crop domestication (Chen, 2007; Soltis et al., 2014). This may explain why the majority of crop plants are of polyploid origin, and many other crops are grown as hybrids. The most widely-cultivated allotetraploid cotton (Gossypium hirsutum L.) is an excellent example of polyploidization and domestication from small trees into an annual crop (Wendel and Cronn, 2003; Guan et al., 2014).
doubling is predicted to have occurred at 1-2 million years ago (MYA) through hybridization and chromosome doubling between the extant progenitors that are closely related to *G. arboreum* (*A_2*) or *G. herbaceum* (*A_1*), originating in the Old World, and *G. raimondii* (*D_5*), originating in the New World (Wendel, 1989; Wendel and Cronn, 2003). Interestingly, the genome size of *G. arboreum* (1,746 Mb/1C) is almost twice the size of *G. raimondii* (885Mb/1C), which is largely related to the expansion of transposable elements (TEs), including long terminal repeat (LTR) retrotransposons (Hawkins et al., 2006; Wang et al., 2010; Zhang et al., 2015). The TE-rich regions often form heterochromatin and are different from genic regions in the euchromatin (Wang et al., 2010; Guan et al., 2014).

Chromatin modifications can be visualized and quantified using immunostaining of metaphase chromosomes that have revealed dramatic regional differences in the distribution of specific histone modifications, particularly in centromeric (constitutive) heterochromatin in plants and animals (Carchilan et al., 2007; Terrenoire et al., 2010). The technique has been applied in meiotic (pachytene) chromosomes in maize to display regional variation of methylated histone isoforms, with distinctive differences between heterochromatic and euchromatic regions (Shi and Dawe, 2006). However, these studies are very limited in plants (Houben and Schubert, 2003; Jin et al., 2008). The differences of histone modifications between homoeologous chromosomes and their consequences on expression of homoeologous genes in a polyploid species are largely unknown.

In this study, distribution patterns of histone modifications (H3K4me3 and H3K27me2) in heterochromatic and euchromatic regions of homoeologous chromosomes in
*G. hirsutum* were quantified and comparatively analyzed with transcriptome data. The results have revealed distinctive patterns of heterochromatin and euchromatin between homoeologous chromosomes and an association of transcriptionally active chromatin with gene-rich regions. Although the overall fluorescent intensities of histone marks are equally distributed between A and D homoeologous chromosomes, the mean fluorescent intensities per chromosome bin of histone modifications linked to active transcription are significantly different between A and D homoeologs. ChIP-seq analysis further shows that different numbers of A- and D-homoeologs with biased H3K4me3 levels among homoeologous pairs in the allotetraploid cotton correlate with the number and levels of expressed genes between A and D homoeologs. Two chromosomes display different expression levels relative to other chromosomes, probably because of translocations. These data indicate that both histone modifications and genome organization contribute to the expression bias of homoeologous genes in allopolyploids.
RESULTS

H3K4me3 distributions among A and D homoeologous chromosomes

To determine histone modification patterns in allotetraploid cotton, metaphase chromosome spreads in root tip cells (Fig. 1A) of *G. hirsutum* were immunostained with antibodies against H3K4me3 (Fig. 1B). Centromeric heterochromatin was consistently unstained, while the arms of most chromosomes showed distinct patterns of strongly and weakly stained regions (Fig. 1C). The signals were enriched at the distal end of all chromosomes in *G. hirsutum* (Fig. 1C; Supplemental Fig. S1A), which were consistent with high levels of expressed genes in the distal regions (Zhang et al., 2015). We distinguished A and D homoeologous chromosomes by genomic DNA *in situ* hybridization (GISH) with genome DNA probe from *G. arboreum* (Supplemental Fig. S1B). Using a combination of chromosome size, chromosome arm ratio, and H3K4me3 signal patterns (Table 1; Supplemental Fig. S1B), we identified all chromosomes and constructed karyotypes based on three reproducible cells, including one shown in Fig. 1D. The order of cytological chromosomes was based on the length of each chromosome, which corresponded to the length of sequenced chromosomes (Table 1). If no significant difference existed between cytological chromosome lengths (±0.05), correlated coefficients were calculated among H3K4me3 immunostaining intensities and gene densities between each chromosome pair, and the highest correlation coefficient value was chosen for chromosome assignment (see below). Overall, the H3K4me3 signal level was nearly equal between A (8,106,887±1,355,268) (total fluorescent luminance intensities per chromosome among all A
chromosomes) and D (8,253,784±1,528,778) homoeologous chromosomes. The A homoeologous chromosomes are twice the size as the D homoeologous chromosomes (Zhang et al., 2015). Thus, signal intensities per chromosome area were higher in the D than in the A homoeologous chromosomes (Fig. 1D; Supplemental Fig. S1B). The mean intensities per area of D chromosomes (61,144) was ~50% higher than the A chromosomes (40,709) (Table 1) (t-test, p= 2.31E-12). The H3K4me3 distribution pattern in the A homoeologs was similar to that in G. arboreum (Supplemental Fig. S2A), the extant A-genome progenitor, suggesting conservation of H3K4me3 distribution patterns from the progenitor to the allotetraploid cotton. The immunostaining patterns of another active marker, histone H4 lysine 12 acetylation (H4K12ac) (Supplemental Fig. S2B), were very similar to those stained
by H3K4me3, except for clustered distributions at nucleolar organizer regions, which may contribute to the differential activity of rDNA loci, a phenomenon known as nucleolar dominance (Pikaard, 1999).

H3K27me2 marks present in nuclear peripherals and devoid in active rDNA loci

H3K27me2 marks correlated with heterochromatin and gene silencing in plants and animals (Martin and Zhang, 2005). Immunostaining using antibodies against H3K27me2 showed that the signals were distributed throughout the metaphase chromosomes of G. arboreum (Supplemental Fig. S3A) and G. hirsutum (Supplemental Fig. S3B). Signals in telomeric regions were enriched in some chromosomes. In the interphase, H3K27me2 antibody signals were clustered as speckles around the nuclear peripherals (Supplemental Fig. S3, C and D). There are three 45S rDNA loci, two major and one minor loci, in G. arboreum (in some cases, minor loci could not be detected by FISH) (Hanson et al., 1995) (Wang et al., 2008). H3K27me2 antibody signals were co-localized with one of the rDNA loci, but not with the other rDNA loci (Supplemental Fig. S3E). This suggests silencing of one rDNA locus and expression of two other loci that are subject to nucleolar dominance (Pikaard, 1999).

Correlation between gene densities and H3K4me3 intensities

Previous studies have identified a complete set of 26 chromosome pairs by BAC-FISH (Wang et al., 2006). Based on that information, we merged the H3K4me3 intensities with the chromosome morphology map in tetraploid cotton (Fig. 1C-D). The combination of morphology and H3K4me3 density was used to assign G. hirsutum chromosomes from A01...
to A13 and from D01 to D13 (Supplemental Fig. S4; Fig. 1). Seven pairs (A01-D01, A04-D04, A05-D05, A06-D06, A08-D08, A09-D09, and A11-D11) had the highest correlation coefficient values (P<0.01); five pairs (A02-D02, A03-D03, A10-D10, A12-D12, and A13-D13) had higher (but not the highest) correlation coefficient values with P<0.01; only one pair (A07-D07) had the value within a low confidence level (Supplemental Table S6). With the chromosomes identified, the maps of antibody intensity and gene density among 13 pairs of A and D homoeologous chromosomes were comparatively analyzed.

*G. hirsutum* genome consisted of similar numbers of genes in A and D homoeologous chromosomes but more repetitive sequences in the A homoeologs (69.1%) than in the D homoeologs (58.0%) (Supplemental Data 1) (Zhang et al., 2015). These repeats were enriched in the centromeres and gradually lost towards distal regions of chromosomes (Fig. 2A). On the contrary, the distribution of genes showed an opposite trend. Because the D homoeologous chromosomes are smaller but essentially with similar numbers of genes (Zhang et al., 2015), the gene density in the D homoeologs should be higher.

In humans, the level of histone modifications (e.g., H3K4me3) is directly proportional to the density of genes along each chromosome (Terrenoire et al., 2010). We tested this relationship in *G. hirsutum* by comparing gene density maps in 5-Mb windows across each chromosome (Supplemental Data 2) with H3K4me3 signal patterns (Table 1). The H3K4me3 levels across each chromosome were quantified by Intensity Profile of the Olympus cellSens Dimension, which could generate the intensity profiles over different positions.
The results indicate that gene density and H3K4me3 intensity maps were consistent in chromosome pairs, such as A05-D05, A09-D09, A11-D11, and A12-D12 (Fig. 2B), except for chromosomes that involve translocations (Menzel et al., 1986; Wang et al., 2006; Zhang et al., 2015) (see below). The chromosome pairs consistent with the gene and H3K4me3
Densities have relatively high gene numbers and evenly distributed gene densities. This is true not only for intensely stained regions (e.g., landmark bands on chromosomes A09q, A05q, D05q and D09q; p and q denote short and long arms of each chromosome), but also for weakly stained areas (e.g., the signal distribution along chromosomal bands A05p and A12q) (Fig. 2B).

For example, chromosomes A09 and D09 have 75.0 and 51.0-Mb DNA, respectively. Assuming a linear relationship between genomic DNA distribution and physical position on the metaphase chromosome, we divided each chromosome into a number of bins with 5 Mb per bin, resulting in 14 bins on chromosome A09 and 11 bins on chromosome D09. These bins were quantified for their overall fluorescence intensities (Fig. 3A). Correlation coefficients between the gene density and H3K4me3 intensity distributions among bins on each chromosome were 0.894 and 0.967 in A09 and D09, respectively (P < 0.01), suggesting a significant correlation between gene density and H3K4me3 intensity.

Transcriptome analysis of root tip cells in *G. hirsutum* cotton

Different levels of chromatin modifications could result in gene expression differences along the chromosomes. We tested this in tetraploid cotton by RNA-seq analysis using sequence libraries prepared from the root-tip cells, which were the same cells used to prepare for metaphase chromosome immune-staining assays. Three independent mRNA-seq libraries were constructed and sequenced by Illumina paired-end reads. A total of 171 million reads were obtained from three libraries, and 85-88% reads were mapped onto A and D homoeologs of *G. hirsutum* sequence (Zhang et al., 2015), among which ~49%
and 51% reads were mapped onto the A and D homoeologs, respectively (Supplemental Table S1). At the whole genome level, an equal number of A (16,711) and D (17,400) homoeologous genes was expressed in the root tip cells of allotetraploid cotton (Supplemental Table S1; Fig. 3B), consistent with the results from expression analysis in over thirty tissues (Zhang et al., 2015). The overall expression levels (fragments per kilobase per million, FPKM) in the A and D homoeologs were proportional to the number of expressed genes (Supplemental Table S1; Supplemental Table S2; Fig. 3B and C) and also consistent with total H3K4me3 signal intensities in the A and D homoeologous chromosomes (Table 1,
Fig. 2A). For example, transcription levels and H3K4me3 immunostaining intensities on each chromosome were significantly correlated (r= 0.868 for A09 and 0.962 for D09, P<0.01) (Fig. 3A).

**Differential gene expression between A and D homoeologs**

Although the total number of genes did not show expression bias in multiple tissues, as many as 30% of genes can be expressed differently between A and D homoeologs in a given tissue (Zhang et al., 2015). The reason for this is unknown. Since we know where each homoeologous pair is located in the genome sequence, we can associate expression patterns with H3K4me3 status of immuno-staining chromosomal bins in root-tip cells. We found that the H3K4me3 intensities per surface area in most D homoeologous chromosomes, except for A03 vs. D03 and A07 vs. D07, were higher than those in the A homoeologs (Fig. 4A), which could lead to biased expression of the D homoeologs. To further test this relationship, we identified 18,470 homoeologous gene pairs using reciprocal BLAST hits between A and D homoeologs, as previously reported (Zhang et al., 2015) and compared them with ChIP-seq data of H3K4me3 obtained from the same-type of root-tip cells (Table S5, Supplemental Data 3). Consistent with the cytological data, more D homoeologs than the A-homoeologs had high H3K4me3 levels among all 13 homoeologous pairs of chromosomes (Fig. 4B).

At the level of individual genes, we identified 5,361 (29%) homoeologous genes with unequal (biased) expression in the root-tip cells. Among them, 2,452 (13%) were expressed towards A homoeologs (A>D), while 2,909 (16%) were expressed towards D homoeologs.
(D>A) (Fig. 4C). The number of genes with D-homoeolog biased expression was higher than that with A-homoeolog biased expression among most homoeologous pairs, except for A03 vs. D03, A05 vs. D05, A07 vs. D07, and A13 vs. D13 pairs (Supplemental Data 3, Fig. 4D).

Among the genes that were hypermethylated at H3K4me3 in the A homoeologs relative to
the D homoeologs, more than three-fold of A homoeologs were expressed at higher levels than the D homoeologs. Likewise, among the genes that were hypermethylated at H3K4me3 in the D homoeologs relative to the A homoeologs, nearly four-fold of D homoeologs were expressed at higher levels than the A homoeologs (Fig. 4E). Further analysis using pairwise comparison of A and D homoeologs with H3K4me3 levels indicated statistically significant correlation between expression and H3K4me3 levels of A and D homoeologs ($r=0.412$, $P < 0.01$, Fig. 5A). Among hypermethylated H3K4me3 loci, more D homoeologs (540) than A homoeologs (332) displayed biased expression ($P<0.01$, using Chi-square test). This biased
expression was also observed for individual homoeologous chromosome pairs. For example, more D homoeologs (49) than A homoeologs (31) on A12 and D12 homoeologous chromosomes showed biased expression (P<0.01, using Chi-square test, Fig. 5B). This trend was observed at the level of several individual genes examined (Fig. 5C). For example, higher levels of H3K4me3 in Gh_A09G0916 and Gh_A12G0383 correlated with higher levels of expression in these loci relative to their respective homoeologs (Gh_D09G0946 and Gh_D12G0281). Likewise, H3K4me3 and expression levels were higher in the D homoeologs (Gh_D09G2159 and Gh_D12G2596) than in their corresponding A homoeologs (Gh_A09G1954 and Gh_A12G2468). Together, these data suggest that H3K4me3 modifications play an important role in the expression bias of homoeologous genes. Overall, more D homoeologs were hypermethylated at H3K4me3 and expressed at higher levels than A homoeologs.

Effects of chromosomal translocation on homoeologous gene expression

Notably, several homoeologous chromosome pairs with equal number of expression-biased genes involved chromosomal translocations (Fig. 4D). Two translocations have occurred between chromosomes A02 vs. A03 and A04 vs. A05 based on cytological and sequencing analyses (Menzel et al., 1986; Wang et al., 2006; Zhang et al., 2015) (Fig. 6A). As a result, the comparison between putative homoeologous pairs did not fully reflect the expected homoeologous relationship prior to translocation. Expression levels of the genes between the translocated regions were different from that of all chromosomes (Fig. 6B). This was consistent with the difference of H3K4me3 distribution levels between
translocated regions and all chromosomes (Fig. 6C). While the number of expression-biased genes of A02-D02 and A03-D03 gene pairs matched the rest of the genome, the number of genes with D-biased expression was significantly higher in the translocated A02-D03 and the A03-D02 regions (Fig. 6D, chi square test, p<0.05). The trend was similar but to a lesser degree in the comparison between A04-D04 vs. A05-D05 and A04-D05 vs. A05-D04. These translocations could affect the local chromatin structure and hence expression levels for the genes involved.
Correlation of homoeolog-biased expression with H3K4me3 could lead to functional consequences (Supplemental Data 4). Gene Ontology (GO) analysis showed enrichments of biased expression of A homoeologous genes in plant cell wall and those of biased expression of D homoeologous genes in metabolic process in root-tip cells (Supplemental Fig. S6). In other tissues such as cotton ovules and fibers, A homoeologous genes with biased expression include fiber-related transcription factor genes, while biased expression of D homoeologs is related to stress responses (Zhang et al., 2015).
DISCUSSION

Euchromatic and heterochromatic marks in plant chromosomes

Histone modifications mediate epigenetic regulation of gene expression, growth, and development in plants and animals (Berger, 2007; Li et al., 2007; Zhang, 2008). Histone marks such as H3K4me3 and H4K12ac are associated with euchromatin and gene activation, while marks like H3K27me2 correlate with heterochromatin and gene repression (Jenuwein and Allis, 2001; Li et al., 2007). In allotetraploid cotton, the H3K4me3 density difference between A and D homoeologous chromosomes could be related to the genome size, which is twice as large in the former as in the latter. In addition, retrotransposons including long terminal repeat (LTR) retrotransposons are more abundant in the A homoeologs compared to D homoeologs (Zhang et al., 2015). Together, more repetitive sequences and larger size in the A subgenome than in the D subgenome are associated with more heterochromatin per megabase in the A homoeologs than in the D homoeologs and consequently, less euchromatin per megabase in the A homoeologs than in the D homoeologs. On each chromosome, the distal portions were enriched with H3K4me3 and H4K12ac so the euchromatin is more abundant near chromosome ends than proximal regions, which is consistent with the distribution of genes and genetic recombination along cotton chromosomes (Paterson et al., 2012). This gradient distribution of H3K4me3 across metaphase chromosomes of *G. hirsutum* is consistent with the notion that H3K4me3 is conserved euchromatic mark in plant species (Fuchs et al., 2006). H4K12ac has very similar distribution patterns as H3K4me3, except for clustering of H4K12ac near the nucleolar
organizing regions (NORs), as previously reported in other plant chromosomes (Belyaev et al., 1997; Jasencakova et al., 2000). This indicates a role for H4K12ac in differential expression of uniparental rRNA genes subjected to nucleolar dominance (Lawrence et al., 2004).

H3K27 methylation has been linked to several silencing phenomena in animals, including X chromosome inactivation and genomic imprinting (Martin and Zhang, 2005). H3K27me2 is highly concentrated in the centromeric heterochromatin in A. thaliana (Lindroth et al., 2004; Mathieu et al., 2005) and in maize (Shi and Dawe, 2006). However, this distribution pattern is not conserved in cotton compared to maize or Arabidopsis (Fuchs et al., 2006). H3K27me2 signals are patchy in the metaphase and display some weak bands in interphase cells (Supplemental Fig. 3). Allotetraploid cotton has two major and one minor NORs (Hanson et al., 1995), and H3K27me2 is more prominent in one of them. This suggests that H3K27me2 could cause silencing of one rDNA locus, while H4K12ac could be associated with active rDNA loci.

Both chromatin and genome organization mediate homoeologous gene expression

After polyploidization, genetic changes such as mutations, sequence elimination, and chromosomal rearrangements, as well as epigenetic mechanisms at transcriptional and post-transcriptional levels can contribute to gene expression variation (Chen, 2007; Doyle et al., 2008; Soltis et al., 2014). Polyploidy may induce epigenetic modifications of homoeologous chromosomes to reprogram gene expression and developmental patterns of allotriploids (Song and Chen, 2015), leading to genome-wide dosage-dependent and
independent gene expression novelty (Shi et al., 2015). Histone marks such as H3K4me3 are known to correlate with gene expression diversity between species and in Arabidopsis allopolyploids (Ha et al., 2011; Shi et al., 2015). In this study, we associated H3K4me3 levels with expression changes of homoeologous genes in allotetraploid cotton. Although the total H3K4me3 levels are relatively equal between A and D homoeologous chromosomes, the fluorescent intensity per area of H3K4me3 is significantly higher in the D homoeologous chromosomes than in the A homoeologs. Consistent with the cytological data, ChIP-seq analysis showed more D homoeologs with biased H3K4me3 levels than A homoeologs with biased H3K4me3 levels, which could loosen chromatin structure for the D homoeologs allowing genes to be expressed at higher levels. Indeed, there are more genes with D-homoeolog biased expression than with A-homoeolog biased expression in most homoeologous chromosome pairs. This suggests an overall more effect of H3K4me3 on D homoeologs than on A homoeologs.

Chromosomal translocations can result in hitherto unforeseen and large-scale changes in gene expression that are the consequence of alterations in normal chromosome territory positioning in human (Harewood et al., 2010). In allotetraploid cotton, homoeologous chromosome pairs with chromosomal translocations could possess different chromatin structure and change the number of genes with homoeolog-biased expression among the chromosomes involved. Expression levels of the genes between the translocated regions were different from that of all chromosomes and were consistent with the difference of H3K4me3 distribution levels between translocated regions and all chromosomes. Thus,
during polyploidization genome reorganization including translocation may alter chromatin structure that induces gene expression changes and potentially phenotypic variation in polyploids. It is difficult to determine the exact relationship between an immunostained and sequenced chromosome unless a set of DNA markers specific to each chromosome is used on the same cells where the chromosomes are immunostained. This would be technically challenging and cannot be resolved at the present stage. The current approach using the chromosome length, immunolabeling intensity, and gene density maps has shown to work well in cotton as in mammalian cells (Terrenoire et al., 2010). The ChIP-seq results have confirmed the cytological data. Distributions between H3K4me3 intensities and gene densities are comparable and consistent between most homoeologous chromosome pairs. With improvement of technologies, chromatin and expression analysis of single cells or chromosomes would increase the resolution and precision of chromatin regulation. This would be extremely valuable to resolve gene expression changes between homoeologous chromosomes in polyploid species for the genetic improvement of important polyploid crops such as wheat (food), cotton (fiber), and canola (oil).

MATERIALS AND METHODS

Immunostaining and FISH

Root tips were harvested from rapidly growing roots of A-genome extant progenitor (Gossypium arboreum L cv. JLZM) and tetraploid cotton (G. hirsutum L acc. TM-1) and treated with N2O for 1-3 hours and fixed in a phosphate buffer solution (PBS), containing 4%
paraformaldehyde. After washing with PBS, the fixed root tips were digested with 4% Cellulase Onozuka R-10 (Yakult Honsha, Tokyo, Japan) and 1% Pectinase Y-23 (MP Biomedicals, Santa Ana, California) for 2 h at 37°C. The digested root tips were washed in PBS and water (10 min each) and squashed in a drop of water. The slides were rapidly deep-freezed with liquid nitrogen. After removing coverslips, slides were immediately treated with PBS, containing 3% BSA and 1% Triton X-100. Commercial antibodies against H4K12Ac (Abcam, Cambridge, Massachusetts) and against H3K4me3, H3K27me2 and H3K9me2 (Upstate Biotechnology, Lake Placid, New York) were diluted into 1:100 in PBS, applied to the slides that were subsequently covered with coverslips. The slides were incubated in a moist chamber at 37°C for overnight. After washing in PBS, the antibodies were detected using the secondary antibodies of Cy3-conjugated anti-rabbit. DNA was counterstained with 4',6-diamidino-2-phenylindole (DAPI). Following histone immunostaining, fluorescence in situ hybridization (FISH) was performed on chromosome spreads according a previously published protocol (Han et al., 2009). Digoxigenin-labeled probes were detected using fluorescein-conjugated anti-digoxigenin (Roche Diagnostics, Indianapolis, Indiana). Chromosomes were counterstained with Vectashield Mounting medium containing DAPI (Vector Laboratories, Burlingame, CA) under a coverslip. To help chromosome identification, we used telomere and genomic in situ hybridization (GISH) for chromosome analysis. Telomeric DNA (42 copies of TTAGGG) was used as digoxigenin-labeled probe (fluorescein-conjugated anti-digoxigenin) for FISH analysis, and
genomic DNA from *G. arboreum* was used as digoxigenin-labeled probe (fluorescein-conjugated anti-digoxigenin) for GISH analysis.

**Analysis of chromosome images and immunostaining signal intensities**

Slides were examined using an Olympus BX53 fluorescence microscope (objective UPlanSApo 100× 1.40, oil) at room temperature (Olympus, Tokyo, Japan). The images of chromosomes, immunostaining and FISH signal channels were captured and merged using Olympus cellSens Dimension software V5.1 with an Olympus DP72 CCD camera. To determine H3K4me3 signal levels of A and D chromosomes, only the cells with non-overlapping chromosomes were analyzed, and the sum of relative signal levels of each chromosome was calculated using Olympus cellSens Dimension. Final image adjustments were processed using Photoshop CS 3.0 (Adobe, San Jose, California). For each chromosome, the level of H3K4me3 intensities was estimated using the Intensity Profile of the Olympus cellSens, which measured the signal intensity per 0.0339 μm per unit. The chromosomes were first partitioned into homologous pairs using morphological features (arm length and short/long arm ratios) and H3K4me3 patterns and assigned based on arm length of cytological chromosome (from short to long). For quantitative analysis, we divided each chromosome into bins (5 Mb per bin) using H3K4me3 fluorescent intensities and gene densities. The correlation between H3K4me3 fluorescent intensities and gene densities in each bin on each chromosome pair was estimated using Pearson correlation coefficients and tested for statistical significance (Supplemental Table S3; Supplemental Table S4). The confidence of homoeology assignment was based on Pearson correlation coefficients and
P-values of immunostaining densities (H3K4me3) at distal regions between A and D homoeologous chromosomes (Supplemental Table S6).

RNA extractions, RNA-seq libraries and sequencing

RNA samples of root tips in three replicates were extracted using a modified mirvana mirna isolation kit procedure (Ambion, Austin, USA), and their quality was evaluated on Nanodrop 2000 (Thermo Fisher Scientific, Waltham, Massachusetts) and Agilent Bioanalyzer (Agilent Technologies, Santa Clara, California). mRNA-seq libraries were constructed following the RNA-seq library prep kit protocol and sequenced by Illumina HiSeq 2500 machine (Illumina, San Diego, California) with 100 base pair-end reads at Nanjing Agricultural University.

RNA-seq data pre-processing and quality control

Raw reads were examined and filtered by the following steps: (1) removing the last index of original FASTQ format with perl script, (2) performing quality control of the sequencing data by NGS QC Toolkit V2.3 (Patel and Jain 2012), and (3) discarding the low quality reads to ensure that more than 80% of the bases of each retained read had the Illumina quality > 20 (q20 indicating 1% sequencing error rate).

Mapping and analysis of gene expression in homoeologous chromosomes

Clean reads were mapped onto the G. hirsutum genome (Zhang et al., 2015). The mapping was processed by TopHat version 2.0.12 (http://ccb.jhu.edu/software/tophat/index.shtml) utilizing Bowtie2 version 2.23 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml). Mapping results generated by
TopHat were filtered to retain only uniquely mapped reads before being transferred into Cuffdiff for estimating read counts. For each gene, the value of fragments per kilobase per million reads (FPKM) was calculated by in-house scripts based on the count table of Cuffdiff output. Only the genes with mean FPKM of ≥1 were used for further analysis. The reads were used to estimate the number of the genes and expression levels of the each homoeolog or chromosome. The same chromosome coordinates were used for visualization and distribution of repeats, genes, and FPKM values in the *G. hirsutum* genome (Zhang et al., 2015) using the program Circos (Krzywinski et al., 2009). Expression bias of homoeologous genes was estimated to identify the genes with differential expression between A and D homoeologs using statistical tests (FDR<0.05) with at least 1.5-fold changes in average expression levels (FPKM) across three biological replicates. The RNA-seq data were deposited at GEO (http://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE71418.

**Chromatin immunoprecipitation (ChIP) and ChIP-seq library preparation**

ChIP was performed by immunoprecipitating chromatin with antibodies against H3K4me3 (Upstate Biotechnology, Lake Placid, New York) with two biological replicates of root tips, as described previously (Saleh et al., 2008). ChIP-seq libraries for immuno-precipitated and input samples from two biological replicates of root tips were constructed and sequenced by Illumina HiSeq 2500, following the manufacturer’s recommendations (Illumina, San Diego, California). The ChIP-seq data has been deposited at GEO (http://www.ncbi.nlm.nih.gov/geo/) with the accession number GSE83371.
Computational analysis of ChIP-seq data

Sequencing reads were mapped onto the *G. hirsutum* reference genome (Zhang et al., 2015) by SOAP2 version 2.21t software using default parameters (Li et al., 2009). MACS software was used to identify peaks (P<0.01) (Zhang et al., 2008). H3K4me3 distributions of homoeologous genes were estimated to identify the genes with differential H3K4me3 levels between A and D homoeologs using statistical tests (FDR<0.05) with at least 1.5-fold changes (FPKM).
Table 1. Chromosome morphology and H3K4me3 intensities in allotetraploid cotton

<table>
<thead>
<tr>
<th>Physical chromosome</th>
<th>Relative length (%) a</th>
<th>Arm ratio (A)</th>
<th>Physical length (bp)</th>
<th>Total signal b</th>
<th>Signal percentage (%) c</th>
<th>Area (µm²)</th>
<th>Intensity per area d</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>A01</td>
<td>4.66±0.10</td>
<td>1.43±0.15</td>
<td>9.99E+07</td>
<td>657,674±162,407</td>
<td>4.02%±0.84%</td>
<td>16.52±2.06</td>
<td>39,923±3,677</td>
<td></td>
</tr>
<tr>
<td>A02</td>
<td>4.43±0.19</td>
<td>1.48±0.15</td>
<td>8.34E+07</td>
<td>466,557±145,342</td>
<td>2.81%±0.44%</td>
<td>15.72±3.28</td>
<td>29,611±2,000</td>
<td></td>
</tr>
<tr>
<td>A03</td>
<td>4.82±0.11</td>
<td>1.35±0.14</td>
<td>1.00E+08</td>
<td>557,364±145,175</td>
<td>3.37%±0.33%</td>
<td>14.91±3.46</td>
<td>38,735±3,474</td>
<td></td>
</tr>
<tr>
<td>A04</td>
<td>3.65±0.17</td>
<td>1.40±0.13</td>
<td>6.29E+07</td>
<td>290,055±96,335</td>
<td>1.74%±0.29%</td>
<td>11.95±1.82</td>
<td>24,189±3,418</td>
<td></td>
</tr>
<tr>
<td>A05</td>
<td>4.57±0.12</td>
<td>1.35±0.18</td>
<td>9.20E+07</td>
<td>1,054,724±249,696</td>
<td>6.4%±0.50%</td>
<td>16.75±1.36</td>
<td>63,725±6,371</td>
<td></td>
</tr>
<tr>
<td>A06</td>
<td>5.13±0.19</td>
<td>2.08±0.10</td>
<td>1.03E+08</td>
<td>589,725±75,662</td>
<td>3.62%±0.28%</td>
<td>15.94±1.59</td>
<td>37,641±3,572</td>
<td></td>
</tr>
<tr>
<td>A07</td>
<td>4.34±0.26</td>
<td>1.46±0.13</td>
<td>7.83E+07</td>
<td>655,358±94,698</td>
<td>4.01%±0.07%</td>
<td>15.53±5.31</td>
<td>41,775±5,902</td>
<td></td>
</tr>
<tr>
<td>A08</td>
<td>5.17±0.20</td>
<td>1.52±0.12</td>
<td>1.04E+08</td>
<td>466,423±91,853</td>
<td>2.84%±0.14%</td>
<td>15.66±2.79</td>
<td>29,889±1,446</td>
<td></td>
</tr>
<tr>
<td>A09</td>
<td>3.76±0.15</td>
<td>2.53±0.12</td>
<td>7.50E+07</td>
<td>673,873±132,876</td>
<td>4.12%±0.55%</td>
<td>13.41±0.73</td>
<td>50,653±4,187</td>
<td></td>
</tr>
<tr>
<td>A10</td>
<td>4.98±0.17</td>
<td>1.34±0.14</td>
<td>1.01E+08</td>
<td>691,678±80,475</td>
<td>4.25%±0.17%</td>
<td>15.05±3.10</td>
<td>46,905±3,083</td>
<td></td>
</tr>
<tr>
<td>A11</td>
<td>4.59±0.28</td>
<td>1.53±0.11</td>
<td>9.33E+07</td>
<td>776,202±81,460</td>
<td>4.77%±0.30%</td>
<td>16.17±2.15</td>
<td>48,862±3,418</td>
<td></td>
</tr>
<tr>
<td>A12</td>
<td>4.51±0.17</td>
<td>2.06±0.29</td>
<td>8.75E+07</td>
<td>847,605±40,103</td>
<td>5.24%±0.55%</td>
<td>16.89±2.08</td>
<td>50,595±2,477</td>
<td></td>
</tr>
<tr>
<td>A13</td>
<td>4.37±0.16</td>
<td>1.52±0.11</td>
<td>8.00E+07</td>
<td>378,649±70,493</td>
<td>2.31%±0.07%</td>
<td>14.63±1.23</td>
<td>26,298±2,094</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td></td>
<td></td>
<td>8,106,887±1,355,268</td>
<td>49.51%±0.55%</td>
<td>199.14±25.82</td>
<td>40,755±3,892</td>
<td></td>
</tr>
<tr>
<td>D01</td>
<td>3.05±0.06</td>
<td>1.20±0.21</td>
<td>6.15E+07</td>
<td>457,157±38,618</td>
<td>2.63%±0.17%</td>
<td>9.38±1.23</td>
<td>50,545±3,492</td>
<td></td>
</tr>
<tr>
<td>D02</td>
<td>3.80±0.08</td>
<td>1.33±0.16</td>
<td>6.73E+07</td>
<td>543,146±193,935</td>
<td>3.27%±0.88%</td>
<td>10.48±1.49</td>
<td>52,736±7,173</td>
<td></td>
</tr>
<tr>
<td>D03</td>
<td>2.49±0.08</td>
<td>1.76±0.05</td>
<td>4.67E+07</td>
<td>424,859±69,337</td>
<td>2.63%±0.52%</td>
<td>9.29±1.30</td>
<td>45,909±2,105</td>
<td></td>
</tr>
<tr>
<td>D04</td>
<td>2.81±0.21</td>
<td>2.05±0.16</td>
<td>5.15E+07</td>
<td>742,481±138,203</td>
<td>4.58%±0.92%</td>
<td>9.35±0.61</td>
<td>80,571±7,593</td>
<td></td>
</tr>
<tr>
<td>D05</td>
<td>3.15±0.15</td>
<td>2.14±0.11</td>
<td>6.19E+07</td>
<td>1,114,191±131,179</td>
<td>6.86%±0.66%</td>
<td>11.23±0.15</td>
<td>99,588±4,848</td>
<td></td>
</tr>
<tr>
<td>D06</td>
<td>3.18±0.08</td>
<td>1.23±0.11</td>
<td>6.43E+07</td>
<td>550,891±34,863</td>
<td>3.40%±0.36%</td>
<td>9.87±0.64</td>
<td>55,963±2,085</td>
<td></td>
</tr>
<tr>
<td>D07</td>
<td>2.86±0.13</td>
<td>1.31±0.06</td>
<td>5.33E+07</td>
<td>584,611±108,816</td>
<td>3.57%±0.14%</td>
<td>11.82±1.61</td>
<td>50,195±3,894</td>
<td></td>
</tr>
<tr>
<td>D08</td>
<td>3.38±0.13</td>
<td>1.29±0.25</td>
<td>6.59E+07</td>
<td>603,040±216,770</td>
<td>3.61%±0.75%</td>
<td>9.93±1.31</td>
<td>59,857±5,890</td>
<td></td>
</tr>
<tr>
<td>D09</td>
<td>2.61±0.09</td>
<td>1.49±0.14</td>
<td>5.10E+07</td>
<td>618,220±109,011</td>
<td>3.77%±0.16%</td>
<td>9.22±0.68</td>
<td>67,618±5,151</td>
<td></td>
</tr>
<tr>
<td>D10</td>
<td>3.17±0.11</td>
<td>1.31±0.17</td>
<td>6.34E+07</td>
<td>560,210±129,677</td>
<td>3.04%±0.47%</td>
<td>9.71±0.94</td>
<td>57,695±3,669</td>
<td></td>
</tr>
<tr>
<td>D11</td>
<td>3.78±0.10</td>
<td>1.79±0.13</td>
<td>6.61E+07</td>
<td>968,916±163,002</td>
<td>5.93%±0.39%</td>
<td>12.40±1.61</td>
<td>78,484±4,944</td>
<td></td>
</tr>
<tr>
<td>D12</td>
<td>2.90±0.16</td>
<td>2.29±0.27</td>
<td>5.91E+07</td>
<td>688,976±140,019</td>
<td>4.22%±0.66%</td>
<td>10.63±1.62</td>
<td>65,350±2,746</td>
<td></td>
</tr>
<tr>
<td>D13</td>
<td>3.04±0.18</td>
<td>1.23±0.17</td>
<td>6.05E+07</td>
<td>485,345±55,347</td>
<td>2.99%±0.31%</td>
<td>11.67±1.36</td>
<td>41,733±1,419</td>
<td></td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td></td>
<td></td>
<td></td>
<td>8,253,784±1,528,778</td>
<td>50.49%±0.55%</td>
<td>134.99±13.03</td>
<td>61,350±5,668</td>
<td></td>
</tr>
</tbody>
</table>

a, Relative chromosomal length  
b, H3K4me3 fluorescence signals across chromosome surface  
c, Percent of fluorescence signal of a single chromosome relative to all A and D chromosomes  
d, Total signal levels per chromosome surface area.  
e, Mean fluorescent intensity per area in the At and Dt.
Supplemental Materials

Supplemental Figure S1. Antibodies against H3K4me3 Immunostaining and telomere locations of metaphase chromosomes in allotetraploid cotton.

Supplemental Figure S2. Immunostaining images generated using antibodies against H3K4me3 and H4K12ac, respectively, in metaphase chromosomes of *G. arboreum*.

Supplemental Figure S3. Immunostaining images of interphase and metaphase chromosomes with antibodies against H3K27me2.

Supplemental Figure S4. Epigenetic chromosome map using antibodies against H3K4me3.

Supplemental Figure S5. Gene Ontology (GO) analysis of the genes with homoeolog expression bias.

Supplemental Table S1. Number and percentage of sequencing reads mapped onto the allotetraploid cotton (AD) genome sequence.

Supplemental Table S2. Number and expression levels of A and D homoeologs.

Supplemental Table S3. Correlation coefficients between immunostaining and gene densities of homoeologous A chromosomes in *G. hirsutum* (TM-1).

Supplemental Table S4. Correlation coefficients between immunostaining and gene densities of homoeologous D chromosomes in *G. hirsutum* (TM-1).
Supplemental Table S5. Number and percentage of sequencing reads of ChIP-seq mapped onto the allotetraploid cotton (AD) genome sequence.

Supplemental Table S6. Correlation coefficients of immunostaining densities (enriched with H3K4me3) in the distal portions between A and D homoeologous chromosomes in G. hirsutum (TM-1).

Supplemental Data 1. Gene densities, repeats, and expression levels in allotetraploid cotton

Supplemental Data 2. H3K4me3 immunostaining levels along each chromosome.

Supplemental Data 3. Levels of H3K4me3 and expression bias (A>D and A<D) of homoeologous genes.

Supplemental Figure S1. Antibodies against H3K4me3 Immunostaining and telomere locations of metaphase chromosomes in allotetraploid cotton. A and D homoeologous chromosomes were discriminated by genomic DNA in situ hybridization (GISH) using total genomic DNA of G. arboreum as a probe. A, Distribution of H3K4me3 and telomere across G. hirsutum metaphase chromosomes. Blue, red, and green indicate chromosomes
(counterstained with DAPI), H3k4me3 (cy3), and telomeres (digoxigenin-labeled probes), respectively; bar=5μm. B, Discrimination between A and D homoeologous chromosomes by genomic in situ hybridization (GISH). Genomic DNA from G. arboreum was used as digoxigenin-labeled probe (fluorescein-conjugated anti-digoxigenin) in GISH analysis. Blue, red, and green signals indicate chromosomes, H3k4me3, A homoeologous chromosomes, respectively. C, A and D homoeologous chromosomes are arranged based on GISH, telomere, and immuno-staining patterns; bar=5μm.

**Supplemental Figure S2.** Immunostaining images generated using antibodies against H3K4me3 and H4K12ac, respectively, in metaphase chromosomes of G. arboreum. A, Immunolabeling images using antibodies against H3K4me3 to metaphase of G. arboreum. Metaphase chromosomes counterstained with DAPI (blue) and antibodies against H3K4me3 (Cy3, red); bar=5μm. B, Immunofluorescence images using antibodies against H4K12ac to metaphase of G. arboreum. Metaphase chromosomes counterstained with DAPI (blue) and antibodies against H3K4ac12 (Cy3, red); bar=5μm.

**Supplemental Figure S3.** Immunostaining images of interphase and metaphase chromosomes with antibodies against H3K27me2. A, Metaphase images of G. arboreum, showing chromosomes (blue, DAPI) and H3K27me2 (red, cy3); bar=5μm. B, Metaphase images of G. hirsutum, showing chromosomes (blue, DAPI) and H3K27me2 (red, cy3); bar=5μm. C, Interphase images of G. arboreum, showing chromosomes (blue, DAPI) and
H3K27me2 (red, cy3); bar=5μm. D, Interphase images of G. hirsutum, showing interphase cell (left, DAPI) and H3K27me2(middle, cy3); bar=5μm. E, The relationship between H3K27me2 and 45S rDNA in G. arboreum, showing chromosomes (blue, DAPI), H3k27me2 (red, Cy3), and 45S rDNA (green, digoxigenin-labeled probes were detected using fluorescein-conjugated anti-digoxigenin); bar=5μm.

Supplemental Figure S4. Epigenetic chromosome map using antibodies against H3K4me3. Histograms indicate gene densities (Y-axis) relative to fluorescent levels using antibodies against H3K4me3 (opposite Y-axis). Lines and unfilled boxes, respectively, indicate distribution patterns of the genes and anti-H3K4me3 fluorescent levels in 5-Mb windows. P-values indicate significant levels for Pearson correlation coefficient (r) between each comparison. The cytological chromosomes are from Figure. 1, and H3K4me3 intensity values are shown in supplemental Data 2.

Supplemental Figure S5. Gene Ontology (GO) analysis of the genes with homoeolog expression bias. Blue and grey histograms indicate enrichment of the genes with A-homoeolog expression bias (A>D) and D-homoeolog expression bias (D>A), respectively.

Supplemental Table S1. Number and percentage of sequencing reads mapped onto the allotetraploid cotton (AD) genome sequence.
Supplemental Table S2. Number and expression levels of A and D homoeologs.

Supplemental Table S3. Correlation coefficients between immunostaining and gene densities of homoeologous A chromosomes in G. hirsutum (TM-1).

Supplemental Table S4. Correlation coefficients between immunostaining and gene densities of homoeologous D chromosomes in G. hirsutum (TM-1).

Supplemental Table S5. Number and percentage of sequencing reads of ChIP-seq mapped onto the allotetraploid cotton (AD) genome sequence.

Supplemental Table S6. Correlation coefficients of immunostaining densities (enriched with H3K4me3) in the distal portions between A and D homoeologous chromosomes in G. hirsutum (TM-1).

Supplemental Data 1. Gene densities, repeats, and expression levels in allotetraploid cotton

Supplemental Data 2. H3K4me3 immunostaining levels along each chromosome.

Supplemental Data 3. Levels of H3K4me3 and expression bias (A>D and A<D) of homoeologous genes.
**Supplemental Data 4.** Correlation of homoeolog-biased expression with H3K4me3 levels

**ACKNOWLEDGMENTS**

We thank Dr. Yan Hu for coordinating sequencing activities and Ms. Jingya Yuan for assistance in fluorescence *in situ* hybridization.

**Figure Legends**

**Figure 1.** Distinctive immunostaining patterns of antibodies against H3K4me3 fluorescent labels across metaphase chromosomes in allotetraploid cotton. A-C, Metaphase chromosomes from *G. hirsutum* were counterstained with 4′,6-diamidino-2-phenylindole (DAPI, pseudocolored red) (A) and immunostained with antibodies against H3K4me3 (Cy3, pseudocolored green) (B), and merged (C). D, Immunostained karyotypes constructed from the chromosome spread shown in A-C, Bar = 5 μm.

**Figure 2.** Correlation of gene density with H3K4me3 levels of A and D homoeologous chromosomes in allotetraploid cotton. A, (a-d) The non-overlapping 500-kb windows indicate (a) transcript density (FPKM) of the genes expressed in the root tip cells from high (tall bars) to low (short bars), (b) density of H3K4me3 in the root tip cells from high (dark green) to low (light green), (c) density of the genes expressed in the root tip cells from high (dark blue) to low (light blue), and (d) density of repeats from high (dark grey) to low (light grey). B, H3K4me3 levels are associated with gene-rich regions among homoeologous
chromosomes in allotetraploid cotton. Examples are A05-D05, A09-D09, A11-D11, and A12-D12 homoeologous chromosomes. Histograms showing gene numbers (Y-axis) relative to 5-Mb windows (X-axis) in each chromosome (see Figure 1).

Figure 3. Correlation of gene and transcript densities with H3K4me3 levels between homoeologous chromosomes A09 and D09. A, Quantitative analysis of anti-H3K4me3 fluorescent levels between homoeologous chromosomes A09 and D09 relative to gene densities and transcript levels in root tip cells. Histograms in the left (A09 chromosome) and right (D09 chromosome) indicate gene densities (upper) and transcript levels (lower) relative to anti-H3K4me3 fluorescent levels (Y-axis, the opposite). Black, unfilled box, and line, respectively, indicate distribution patterns of transcripts, genes, and anti-H3K4me3 fluorescent levels in 5-Mb windows. P-values indicate significant levels of Pearson correlation coefficient (r) between each comparison. H3K4me3 intensity values are shown in supplemental Data 2. B, Total number of A and D homoeologous genes expressed in the root-tip cells. C, Total expression levels (FPKM) of the A and D homoeologous genes in the root-tip cells.

Figure 4. H3K4me3 and gene expression level differences between homoeologous chromosomes in the root-tip cells of allotetraploid cotton. A, Anti-H3K4me3 fluorescent intensities in 13 homoeologous chromosome pairs from immunostaining cells. Single (*) and double (**) asterisks indicate statistically significant levels of p<0.05 and p<0.01, respectively,
using Student’s t-test, Error bar ± s.d. B, ChIP-seq data showed the number of genes with
A-homoeolog bias (A>D) and D-homoeolog bias (A<D) of H3K4me3 levels among 13
homoeologous chromosome pairs. Double (**) asterisks indicate statistically significant
levels of p<0.01 using Chi-square test. C, Number and percentage of homoeologs showing
equal (A=D), A-homoeolog biased (A>D), and D homoeolog-biased (A<D) expression. D,
Number of the genes with A-homoeolog biased (A>D) and D-homoeolog biased (A<D)
expression among 13 homoeologous chromosome pairs. Single (*) and double (**) asterisks
indicate statistically significant levels of p<0.05 and p<0.01, respectively, using Chi-square
test. E, Number of genes with A-homoeolog biased (A>D) and D-homoeolog biased (A<D)
expression relative to hyper-H3k4me3 levels in A homoeologs or D homoeologs. Double (**)
asterisks indicate statistically significant levels of p<0.01 using Chi-square test.

Figure 5. Correlation of homoeolog-biased expression with H3K4me3 levels. A, Pairwise
comparison of H3K4me3 levels (Y-axis, D homoeologs; X-axis, A homoeologs) for 5,361 pairs
of homoeologous genes. Purple, orange and grey dots indicate A-biased (A>D), D-biased
(A<D), and equal (A=D or no difference) expression, respectively. P-value of Spearman
correlation coefficient (r) is shown. B, Pairwise comparison of H3K4me3 levels (Y-axis, D
homoeologs; X-axis, A homoeologs) for homoeologous genes from A12 and D12
chromosomes. Dots are shown as in A. P-value of Spearman correlation coefficient (r) is
shown. C, Examples of hyper-H3K4me3 levels correlating with enhanced expression of the A
Figure 6. Chromosomal translocations and their effects on biased gene expression. A,

Diagram of two translocations between chromosomes A02 and A03, and A04 and A05. (a)

Diagram of chromosomes; (b-d) Non-overlapping 500-kb windows indicate (b) repeat
density from high (dark grey) to low (light grey), (c) density of the H3K4me3 levels in the
root-tip cells from high (dark green) to low (light green), and (d) transcript density (FPKM) of
the genes expressed in the root-tip cells. Color ribbons indicate translocation regions. B,

Relative expression levels of the genes among the translocated chromosomes. The gene
expression was divided into four levels (from the lowest to the highest). Expression level 4
(the highest) of homoeologous genes on translocated A chromosomes was shown. The
control was expression level 4 of homoeologous genes among all A chromosomes. An
asterisk indicates statistical significance of p<0.05, using Student’s t-test. C, H3K4me3 levels
of the level 4-genes among the translocated chromosomes. An asterisk indicates statistical
significance of P<0.05, using Student’s t-test. D, Distribution of the genes with A-homoeolog
expression bias (A>D) and D-homoeolog expression bias (A<D) in the translocated regions
between A02 and A03 and between A04 and A05 homoeologous chromosomes. Single (*)
and double (**) asterisks indicate statistically significant levels of p<0.05 and p<0.01,
respectively, using Chi-square test.


Ha M, Ng DW, Li WH, Chen ZJ (2011) Coordinated histone modifications are associated with gene expression variation within and between species. Genome Res 21: 590-598


Shi J, Dawe RK (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. Genetics 173: 1571-1583


