Evolutionary dynamics of chromatin structure and duplicate gene expression in diploid and allopolyploid cotton

Guanjing Hu (huguanjing@gmail.com)
Institute of Cotton Research, Chinese Academy of Agricultural Sciences https://orcid.org/0000-0001-8552-7394

Corrinne Grover
Iowa State University https://orcid.org/0000-0003-3878-5459

Daniel Vera
Florida State University

Pei-Yau Lung
Florida State University

Senthil Girimurugan
Florida Gulf Coast University

Emma Miller
Iowa State University

Justin Conover
University of Arizona

Shujun Ou
Iowa State University https://orcid.org/0000-0001-5938-7180

Xianpeng Xiong
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences

De Zhu
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences

Dongming Li
Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences

Joseph Gallagher
USDA/Agricultural Research Service

Joshua Udall
USDA/Agricultural Research Service

Xin Sui
Florida State University

Jinfeng Zhang
Florida State University

Hank Bass
Florida State University https://orcid.org/0000-0003-0522-0881
Jonathan Wendel
Iowa State University https://orcid.org/0000-0003-2258-5081

Article

Keywords:

Posted Date: October 19th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3373364/v1

Additional Declarations: There is **NO** Competing Interest.

Abstract

Polyploidy is a prominent mechanism of plant speciation and adaptation, yet the mechanistic understandings of duplicated gene regulation remain elusive. Chromatin structure dynamics are suggested to govern gene regulatory control. Here we characterized genome-wide nucleosome organization and chromatin accessibility in allotetraploid cotton, Gossypium hirsutum (AADD, 2n=4X=52), relative to its two diploid parents (AA or DD genome) and their synthetic diploid hybrid (AD), using DNS-seq. The larger A-genome exhibited wider average nucleosome spacing in diploids, and this inter-genomic difference diminished in the allopolyploid but not hybrid. Allopolyploidization also exhibited increased accessibility at promoters genome-wide and synchronized cis-regulatory motifs between subgenomes. A prominent cis-acting control was inferred for chromatin dynamics and demonstrated by transposable element removal from promoters. Linking accessibility to gene expression patterns, we found distinct regulatory effects for hybridization and later allopolyploid stages, including nuanced establishment of homoeolog expression bias and expression level dominance. Histone gene expression and nucleosome organization are coordinated through chromatin accessibility. Our study demonstrates the capability to track high resolution chromatin structure dynamics and reveals their role in the evolution of cis-regulatory landscapes and duplicate gene expression in polyploids, illuminating regulatory ties to subgenomic asymmetry and dominance.

Introduction

Polyploidy is a widespread biological phenomenon in eukaryotes and is important in all levels of biological organization (Fox et al., 2020). Being exceptionally prevalent in ferns and flowering plants (Initiative & One Thousand Plant Transcriptomes Initiative, 2019; Jiao et al., 2011; Ruprecht et al., 2017), whole genome duplications resulting from polyploidy have significant implications for plant physiology, ecology, and evolution (Heslop-Harrison et al., 2022; Leitch & Leitch, 2008; Levin, 1983; Ramsey & Schemske, 2002; Soltis & Soltis, 2016; Stebbins, 1940; Van de Peer et al., 2017, 2021; Wendel, 2015; Wendel et al., 2018). Polyploidy may be associated with expanded ecological ranges (Arrigo et al., 2016; Baniaga et al., 2020; Coughlan et al., 2017; Elliott et al., 2022; Mata et al., 2023; Parshuram et al., 2022; G. Wang et al., 2021; L. Zhao et al., 2022), enhanced tolerance to biotic and abiotic stresses (reviewed in (Van de Peer et al., 2021)), physiological changes (Coate et al., 2012; Knight & Beaulieu, 2008; Mishra, 1997; Orr-Weaver, 2015; Otto, 2007; Sugiyama, 2005), and altered biosynthetic pathways (Combes et al., 2022). These changes may confer economically or ecologically important traits (Heslop-Harrison et al., 2022). Unsurprisingly, numerous vital crop species are relatively young polyploids (Heslop-Harrison et al., 2022; Olsen & Wendel, 2013; Renny-Byfield & Wendel, 2014; K. Zhang et al., 2019).

Increases in whole genome content resulting from polyploidy is often associated with changes in nucleotypic characters, such as cell size, nuclear volume, and cell cycle duration (Doyle & Coate, 2019; Wendel et al., 2018). These genomic changes may also alter epigenetic dynamics, gene expression, the proteome, and molecular networks. One extensively demonstrated effect is the profound rewiring of transcriptomes in response to genomic merger and doubling during allopolyploidization (Giraud et al.,

2021; Grover et al., 2012; Hu & Wendel, 2019; Shan et al., 2020; Visger et al., 2019). This genome-wide rewiring encompasses a diversity of phenomena, including unequal expression of homoeologs at the genic level (referred to as “homoeolog expression bias”) (Flagel et al., 2008; Grover et al., 2012) or the genomic level (“genome dominance”) (Schnable et al., 2011), inconsistency in homoeolog biases across tissues or conditions (“expression subfunctionalization and neofunctionalization”) (Adams et al., 2003) even at the single cell level (K. Zhang et al., 2023), apparent trans-control of duplicate expression (“expression level dominance”) (Grover et al., 2012; Rapp et al., 2009; Yoo et al., 2014; Yoo & Wendel, 2014), and altered co-expression gene networks (Gallagher et al., 2016; Hu et al., 2016). While these studies shed light on the evolutionary dynamics of polyploid transcriptomes, the mechanistic underpinnings of these phenomena remain elusive, limiting our understanding of duplicate gene expression evolution, and hence the origin of evolutionary innovation accompanying polyploidy.

The study of chromatin structure has emerged as a field that may bridge the gap between genome evolution and transcriptome evolution, providing insights into the dynamics of gene expression regulation. The chromatin structure landscape reflects multiple and complex regulatory layers that fine-tune gene expression (Ahmad et al., 2022; Talbert et al., 2019). Nucleosomes, the fundamental structural units of chromatin, consist of 147 bases of DNA wrapped around a core histone octamer (Luger et al., 1997). Facilitating the compaction of genomic DNA into chromatin, nucleosomes play a crucial role in controlling DNA accessibility for processes such as gene transcription, DNA replication, repair, and recombination (Andrews & Luger, 2011; Kornberg, 1974). During transcriptional activation, nucleosomes can be moved to expose or conceal cis-regulatory DNA sites, or transiently destabilized (referred to as “fragile” nucleosomes) at promoter regions (Klemm et al., 2019; Mieczkowski et al., 2016; Zlatanova et al., 2008). Thus, nucleosomes act as regulators of chromatin accessibility, which inherently manifest the myriad epigenetic modifications of histones and DNA that collectively control gene expression (Giles & Taberlay, 2019; Hofmeister et al., 2017; Jackson, 2017; W. T. Jordan & Schmitz, 2016; Kawakatsu et al., 2016; Klein & Hainer, 2020; Niederhuth et al., 2016; Schmitz et al., 2011; Song et al., 2017; Springer & Schmitz, 2017). Understanding the factors that determine nucleosome properties and their impact on chromatin accessibility and gene activity is a central biological challenge.

Over the past decade, high-throughput techniques have been employed in plants to map nucleosome occupancy and chromatin accessibility at a genome-wide scale (Baldi et al., 2020; Barbier et al., 2021; Galli et al., 2020; K. W. Jordan et al., 2020; Liu et al., 2015; Tsompana & Buck, 2014; Voong et al., 2017; Zhang et al., 2015; Zhang & Jiang, 2018; Zhao et al., 2020). These methods, including micrococcal nuclease sequencing (MNase-seq), DNase I hypersensitive site sequencing (DNase-seq), and Assay for Transposase Accessible Chromatin sequencing (ATAC-seq), are based on the physical accessibility of chromatin to nucleases. The nuclease cleavage patterns are used to distinguish accessible DNA regions from nucleosome-protected or transcription factor (TF)-protected regions through fragmentation, tagmentation, or elimination. Since the 1970s, DNase I-hypersensitive sites (DHSs) have been considered a hallmark of active regulatory regions in eukaryotic genomes (Weintraub & Groudine, 1976; Wu et al., 1979; Wu et al., 1979). High-throughput DHS mapping has provided genome-wide insight into cis-regulatory DNA elements (CREs) and TF binding sites (TFBSs) in various plant species (Han et al., 2020,

2022; Jiang, 2015; Qiu et al., 2016; Sullivan et al., 2015; Zhang et al., 2012; Zhao et al., 2018). ATAC-seq, a more efficient alternative to DNase-seq, enables fast and low-input profiling of chromatin accessibility (Lu et al., 2017a), even at the single-cell level (Dorrity et al., 2021). These techniques, along with their variants, have provided insights into cis-regulatory landscapes and gene regulatory networks in plant species (Lu et al., 2019; Reynoso et al., 2022; Ricci et al., 2019).

MNase-seq, on the other hand, is historically used for profiling nucleosome occupancy and has been demonstrated in plants such as Arabidopsis (Chodavarapu et al., 2010; Li et al., 2014; Liu et al., 2015) and rice (Wu et al., 2014; Zhang et al., 2015). Recent applications of this technique utilize two MNase digest conditions, light and heavy, which provides both nucleosome positioning data and chromatin accessibility/sensitivity profiling (Vera et al., 2014; Rodgers-Melnick et al., 2016). That is, differential nuclease sensitivity (DNS) profiling of nucleosome occupancy leads to identifying various levels of chromatin accessibility; this approach was first established in maize based on DNA microarray (Vera et al., 2014), and next employed high-throughput sequencing for genome-wide profiling (Rodgers-Melnick et al., 2016). Like DHS identified by DNase-seq and ATAC-seq, the MNase sensitive footprints (MSFs) from DNS-seq are enriched at the 5’ and 3’ boundaries of genes, and are positively associated with gene expression levels, DNA hypomethylation, conserved noncoding sequences, and known TF binding sites. In maize, MNase hypersensitive regions account for less than 1% of the genome, but are linked to genotypic variants that explain ~40% of variation in phenotypic traits, on a par with coding regions (~48%) (Rodgers-Melnick et al., 2016). Additionally, MNase-profiled cis-regulatory landscapes have been linked to tissue-specific transcription and environmental responses, highlighting their roles in shaping phenotypic variation (Parvathaneni et al., 2020; Pass et al., 2017). A related assay based on small DNA fragments from light MNase digestion, MOA-seq, was recently developed to map small particles that delineate likely TF occupancies at cis-regularoty elements within accessible chromatin regions (Liang et al., 2022; Savadel et al., 2021). Overall, the properties of MNase as a probe for chromatin structure has proven highly informative for characterizing chromatin landscapes, nucleosome positioning, nucleosome stability, and the identification of functional CREs.

The cotton genus, Gossypium, is well-established as a model for the study of evolutionary genomics of polyploidy. More than 50 species are known (Hu et al., 2021; Viot & Wendel, 2023; Wendel & Grover, 2015), and new cotton species continue to be discovered (Gallagher et al., 2017; Stewart et al., 2015). Phylogenetic analyses (Z. Chen et al., 2017; Wendel et al., 2010; Wendel & Cronn, 2003) and genome sequence data (Huang et al., 2021) indicate that the genus originated ~5-10 million years ago. Allopolyploid cottons (AD genome) originated in the Pleistocene following trans-oceanic dispersal of an A-genome progenitor to the New World, where it hybridized with a native D-genome diploid. Allopolyploids subsequently diversified into lineages now represented by seven species, including the commercially important G. hirsutum (Upland cotton) and G. barbadense (Sea Island cotton), each domesticated within the last 7000 years (Wendel & Grover, 2015). The closest extant species related to the D-genome progenitor is G. raimondii, whereas the two A-genome species, G. arboreum and G. herbaceum, are equally good models of the female (seed) parent in the initial hybridization (Wendel et al., 1989). This

well-understood evolutionary history of Gossypium renders it an excellent model for studying allopolyploidy.

Previous studies have highlighted several aspects of duplicate gene expression evolution in Gossypium, including “homoeolog expression bias” (HEB), whereby one of the two homoeologs is more highly expressed than the other, and “expression level dominance” (ELD), an enigmatic phenomenon whereby the total expression of both homoeologs is statistically indistinguishable from the expression level of only one of the two parents (Gallagher et al., 2020; Grover et al., 2012; Hu et al., 2013, 2014, 2015; Rapp et al., 2009; Yoo et al., 2013). Cis and trans regulatory control of expression has also been studied in allopolyploid cotton, with trans regulatory variants preferentially accumulating during about 5000-8000 years of domestication (Bao et al., 2019). These and other regulatory changes in cotton are associated with or causally connected to aspects of the chromatin landscape, including DNA methylation (Song et al., 2017), histone modification (Zheng et al., 2016), chromatin accessibility (Han et al., 2022), and 3D genomic topology (Wang et al., 2018), but to date, the molecular mechanisms underlying chromatin remodeling and its impact on duplicate gene expression remains largely unknown.

Here we applied DNS-seq to comprehensively profile genome-wide chromatin accessibility and nucleosome organization in allopolyploid cotton G. hirsutum, relative to its model diploid progenitors and a synthetic, diploid F1 hybrid that mimics the natural hybridization that occurred 1-2 million years ago (mya). In addition to characterizing the dynamics of chromatin structure change accompanying genomic merger and doubling, we also examined duplicated gene expression patterns to unravel the connections between chromatin remodeling and gene regulation in allopolyploid cotton. Taken together, our study provides a detailed view of the evolutionary dynamics of chromatin structure and cis-regulatory landscapes, highlights how these are altered by genome merger and doubling, and sheds light on their regulatory roles in duplicated gene expression evolution.

Methods

Plant materials

Four Gossypium genotypes were used, including a natural allopolyploid (AD genome), G. hirsutum cultivar Acala Maxxa (AD1), and its model (A- and D- genome) diploid progenitors, i.e., G. arboreum accession A2-101 (A2) and G. raimondii (D5). The two diploid genome groups, A and D, last shared a common ancestor 5-10 mya (Wendel & Albert, 1992), and have diverged to the extent that genome sizes differ two-fold. Thus, the corresponding interspecific diploid F1 hybrid (A2×D5) was included to study the immediate consequences of the merger to two diverged genomes (in the absence of genome doubling and evolutionary time since polyploidization). Four to five plants per genotype were grown in the Bessey Hall Greenhouse at Iowa State University (Ames, Iowa, USA) under controlled short-day conditions (10 hour photoperiod with darkness from 5pm to 7am; 22/28°C, night/day). Mature leaf tissue was harvested from flowering branches at 5 pm, and immediately flash frozen in liquid nitrogen and stored at −80°C.

DNS-seq experiment and data preprocessing

Nuclei isolation. Nuclei were isolated using a modified protocol from Vera et al (2014). Briefly, four grams of frozen tissue were ground together with 10% (w/w) of polyvinylpolypyrrolidone under liquid nitrogen using a mortar and pestle, immediately followed by formaldehyde cross-linking for 10 minutes (min) in 40 mL fixation buffer (1.0 M 2-methyl-2,4-pentanediol, 10 mM PIPES·NaOH at pH 7.0, 10 mM MgCl2, 2% polyvinylpyrrolidone, 10 mM sodium metabisulfite, 5 mM β-mercaptoethanol, 0.5% sodium diethyldithiocarbamate trihydrate, 200 mM L-lysine, and 6 mM EGTA at pH 7.0) containing 1% formaldehyde. Fixation was stopped by adding 2 mL of 2.5 M glycine and stirring for 5 min. To degrade and solubilize organelles, 4 mL of 10% Triton X-100 was added to suspension, followed by stirring for 10 min. The suspension was filtered through one layer of Miracloth (Calbiochem) twice and placed in 50-mL centrifuge tubes. Nuclei were pelleted by centrifugation at 2,000 × g for 15 min at 4°C and subsequently washed three times in 40 mL wash buffer (0.5 M 2-methyl-2,4-pentanediol, 10 mM PIPES·NaOH at pH 7.0, 10 mM MgCl2, 0.5% Triton X-100, 10 mM sodium metabisulfite, 5 mM β-mercaptoethanol, 200 mM L-lysine, and 6 mM EGTA at pH 7.0).

MNase digestion and DNA extraction. Nuclei pellets were resuspended in 2 mL MNase digestion buffer (50 mM HEPES at pH 7.6, 12.5% glycerol, 25 mM KCl, 4 mM MgCl2, and 1 mM CaCl2) and distributed into 500 uL aliquots. Different levels of nuclei digestion were conducted using either 5.6 U/mL (heavy) or 0.4 U/mL (light) MNase, both of which were incubated at 37°C for 10 min. Digestion was stopped by adding 50 mM EGTA on ice for 5 min. Digested nuclei were de-cross-linked at 65°C overnight in the presence of 1% SDS and 100 μg/mL proteinase K, and then treated with 40 μg/mL DNase-free RNaseA at 37 °C for an hour. DNA was extracted by phenol-chloroform extraction and precipitated with ethanol. Extracted DNA was electrophoresed on a 2% agarose gel to inspect the MNase digestion ladders. DNA fragments smaller than 200 bp were purified with the Axygen™ AxyPrep Mag™ PCR Clean-up Kit (Fisher Scientific), following a double-sided SPRI bead size selection (0.9× followed by 1.1×).

Library preparation and sequencing. DNA concentration was measured using the Qubit DNA Assay Kit with a Qubit 2.0 Fluorometer (Life Technology). Sixteen DNA sequencing libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (NEB), according to manufacturer instructions. Indexed libraries were pooled and sequenced on ten Illumina HiSeq 2500 lanes with paired-end 150-cycle sequencing.

Data processing. After quality filtering and trimming of adaptor sequences using CutAdapt (Martin, 2011), paired-end reads generated from the different Gossypium species were mapped against their corresponding reference genomes downloaded from CottonGen (J. Yu et al., 2014), including G. hirsutum cv. TM1 UTX v2.1 (Chen et al., 2020a), G. arboreum cv. SXY1 WHU-updated v1.0 (Huang et al., 2020) and G. raimondii JGI v2.0 (Paterson et al., 2012). The F1 hybrid was mapped against a combined reference of G. arboreum and G. raimondii. Following Bowtie2 [v2.5.1] mapping with options “no-mixed,” “no-discordant,” “no-unal,” and “dovetail” (Langmead & Salzberg, 2012), alignments of quality score ≥20 were retained for following analyses. Based on mapping read coverage, the deepTools [v2.5.2] (Ramírez et al., 2014) commands plotCorrelation and plotPCA were used to assess the reproducibility between replicates and the clustering of different MNase experiments; computeMatrix and plotHeatmap were used to

visualize signal aggregation over genomic regions of interest, e.g. transcription start sites (TSS) and transcription termination sites (TTS). Read coverage data was converted to bigWig files using the UCSC Genome Bioinformatics utility (https://github.com/ucscGenomeBrowser/kent) code "bedGraphToBigWig", and visualized on the Broad Institute Integrative Genomics Viewer (IGV) (J. T. Robinson et al., 2011).

Nucleosome calling, classification and prediction

From the heavy MNase digestion, filtered MNase-seq read alignments were imported in R/Bioconductor framework version 3.5.0 and analyzed using the package nucleR (Flores & Orozco, 2011). Paired-end reads under 260 bp were trimmed to 50 bp around the DNA fragment center. Genome-wide coverage in reads per million (RPM) was computed and normalized using the total number of read alignments from each sample. Noise filtering and peak calling were performed using the following nucleR parameters: pcKeepComp=0.02, peak width=147 bp, peak detection threshold=35%, minimal overlap=50 bp. If the identified peak width is above 150 bp, this peak is considered to contain more than two overlapped nucleosome dyads. Among the non-overlapped nucleosome calls with peak width below 150 bp, well-positioned (W) nucleosomes were defined with peak height score above 0.6 and peak width score above 0.4, while the rest were classified as weakly-positioned, or fuzzy (F) nucleosomes. Nucleosome coverage (NC) is defined as the percentage of genomic regions being occupied by nucleosomes. Nucleosome repeat length (NRL) is defined as the length of DNA wrapped around the histone octamer plus linker DNA, or the center to center distance between consecutive nucleosomes, which were estimated using NucTools scripts "nucleosome_repeat_length.pl" and "plotNRL.R" (Vainshtein et al., 2017). The R package NuPoP (Xi et al., 2010) was used for nucleosome positioning prediction from genomic DNA sequence, which explicitly models the linker DNA length with either a fourth order or first order hidden Markov chain. NuPoP outputs the Viterbi prediction of optimal nucleosome position map, based on which the predicted nucleosome coverage and NRL values were calculated.

Mapping accessible chromatin regions (ACRs) by DNS-seq

MNase sensitive footprints (MSFs). Given the high level of reproducibility (Pearson's $r > 0.9$), mapping results from the two biological and technical replicates per MNase digestion and per genotype were pooled to generate the differential nuclease sensitivity (DNS) profile for each genotype. Using a differential MNase-seq data processing pipeline previously established (Turpin et al., 2018), sequential computation steps were performed to (1) normalize the mapping read coverage in RPM between light and heavy MNase digestions, (2) calculate DNS scores as the difference from light minus heavy read coverages, (3) produce genome-browser-ready data tracks, and (4) identify positive (MNase sensitive) and negative (MNase resistant) peaks using the genomic segmentation algorithm, iSeg [v1.3.4] (Girimurugan et al., 2018). To enable comparisons between species and (sub)genomes, an additional step of quantile normalization was performed before iSeg, normalizing the genome-wide DNS scores across diploid genomes (A2 and D5) and subgenomes (At and Dt) in hybrid and tetraploid cottons. A range of biological cutoff (BC) stringencies were tested in calling the MNase sensitive (MSFs) and

resistant footprints (MRFs), represented by positive and negative DNS peaks, respectively, as previously termed (Vera et al., 2014). An optimized stringency BC=6.0 was used (Supplementary Text 1. Optimization of iSeg stringency) to generate the final list of MSFs.

Sub-nucleosomal particle occupancy (SPO). As previously reported (Grossman et al., 2018; Savadel et al., 2021), small sequence fragments (0-130 bp) from the light MNase digestion can also be used to directly profile the occupancy of sub-nucleosome sized particles involved in transcriptional control. Using awk and BEDTools [v2.27.1] (Quinlan, 2014), the geometric center of each small alignments (0-130 bp) from the light digestion was extracted and intersected with 21 bp sliding genomic windows with a step size of 5 bp. The smoothed profile of small fragment centers was normalized in RPM as the genome-wide SPO scores. Different from the relative scores of DNS, quantile normalization of SPO scores across genomes would lead to substantial signal loss, so the resulting BedGraph files per genome were subjected to iSeg [v1.3.4] separately using optimized stringencies (Supplementary Text 1. Optimization of iSeg stringency). The resulting list of segments represents ACRs identified by SPO.

Mapping ACRs by ATAC-seq and DNase-seq

ATAC-seq. Two replicated ATAC-seq experiments were conducted using the mature leaf tissue of G. raimondii, following a protocol described previously (Lu et al., 2017b). For each replicate, approximately 200 mg freshly collected leaves or flash frozen leaves were immediately chopped with a razor blade in 1 ml of pre-chilled lysis buffer (15 mM Tris-HCl pH 7.5, 20 mM NaCl, 80 mM KCl, 0.5 mM spermine, 5 mM 2-mercaptoethanol, 0.2% Triton X-100). The chopped slurry was filtered twice through miracloth and once through a 40 μm filter. The crude nuclei were stained with DAPI and loaded into a flow cytometer (Beckman Coulter MoFlo XDP). Nuclei were purified by flow sorting and washed in accordance with Lu et al. (Lu et al., 2017b). Sorted nuclei were incubated with 2 μl Tn5 transposase in a 40 μl tagmentation buffer (10 mM TAPS-NaOH ph 8.0, 5 mM MgCl2) at 37°C for 30 minutes without rotation. Integration products were purified using a Qiagen MinElute PCR Purification Kit or NEB Monarch™ DNA Cleanup Kit and then amplified using Phusion DNA polymerase for 10-13 cycles. PCR cycles were determined as described previously (Buenrostro et al., 2013). Amplified libraries were purified with AMPure beads to remove primers. ATAC-seq libraries were sequenced in paired-end 35 bp at the University of Georgia Genomics & Bioinformatics Core using an Illumina NextSeq 500 instrument.

DNase-seq. Public data from cotton young leaves were previously reported (Han et al., 2022; Wang et al., 2017, 2018) and downloaded from NCBI (Supplementary Table S1. Summary of sequencing data in this study).

Data processing. Raw ATAC-seq and DNase-seq reads were adapter and quality trimmed, and then filtered using “Trim Galore” [v0.4.5] (Krueger, 2015). Clean reads were subsequently aligned to corresponding reference genomes using Bowtie2 [v2.3.4] (Langmead & Salzberg, 2012) with the parameters “–no-mixed --no-discordant –no-unal –dovetail”. Three different sets of peak calling methods were tested for ATAC-seq as follows (Supplementary Text 2: ATAC-seq analysis in G. raimondii), and the MACS2 method was used for DNase-seq.

HOMER and MACS2 peak calling. Duplicate reads were removed using Picard [v2.17.0] with default parameters (http://broadinstitute.github.io/picard/). Only uniquely mapped read pairs with a quality score of at least 20 were kept for peak calling. Phantompeakqualtools [v1.14] (Landt et al., 2012) was used to calculate the strand cross-correlation, and deepTools [v2.5.2] (Ramírez et al., 2016) was used to calculate correlation between replicates. The peak calling tool from HOMER [v4.10] (Heinz et al., 2010), i.e., findpeaks, was run in "region" mode and with the minimal distance between peaks set to 150 bp. MACS2 [v2.1.1] (Y. Zhang et al., 2008) callpeak, a second peak-calling algorithm, was run with the parameter "-f BAMPE" to analyze only properly paired alignments, and putative peaks were filtered using default settings and false discovery rate (FDR) < 0.05. Due to the high level of mapping reproducibility by deepTools (Pearson's correlation r = 0.99 and Spearman correlation r = 0.77), peaks were combined and merged between replicates for each tool using BEDTools [v2.27.1] (Quinlan, 2014). BEDTools was also used to intersect HOMER peaks and MACS2 peaks to only retain peak regions identified by both tools as ATAC accessible chromatin regions for subsequent analyses.

Genrich peak calling. Post-alignment steps and peak calling for multiple replicates collectively were performed with one command using Genrich [v0.6.1] (https://github.com/jsh58/Genrich), which was developed and extensively tested in the Harvard FAS Informatics group. The alignment files from both replicates were collectively analyzed by Genrich with the options to remove PCR duplicates (-r), keep unpaired alignments by extending to the average fragment length (-x), exclude problematic genomic regions (-E blacklist.bed), and call peaks using a maximum q-value of 0.05 (-q 0.05) and a minimum AUC of 20.0 (-a 20.0). The output file produced by Genrich is in ENCODE narrowPeak format, listing the genomic coordinates, peak summit, and various statistics for each identified peak.

ACR characterization

Genomic annotation. Various sources of ACRs were identified as described above, including MSFs, SPO regions, and ATAC-seq peaks. An additional filtering step was applied to remove a blacklisted region in G. raimondii (Supplementary Text 3: A hypersensitive region in G. raimondii chromosome 1). According to proximity to the nearest genes, these ACRs were categorized as genic (gACRs; overlapping a gene), proximal (pACRs; within 2 kb of a gene), or distal (dACRs; >2 kb from a gene). To compare GC content between ACRs and non-accessible genomic regions, the BEDTools shuffle command was used to generate the distal (by excluding genic and 2 kb flanking regions) and genic/proximal control regions (by including genic and 2 kb flanking regions), and the nuc command was used to calculate GC content for each ACR and permuted control regions. Using R package ChIPseeker [v1.18.0] (Yu et al., 2015), gACRs and pACRs were combined and further annotated into the following subcategories: promoter (<1 kb, 1-2 kb, 2-3 kb), exon, intron, downstream (<1 kb, 1-2 kb, 2-3 kb), and intergenic regions (>3 kb upstream from TSS and >3 kb downstream from TTS).

Relative to transposable elements (TEs). Whole-genome TE annotation was performed for all reference genomes using the EDTA [v1.9.5] (Ou et al., 2019) pipeline. The proportion of ACR within various TE superfamilies were calculated when the ACR coordinates intersect with a TE interval. Random control

regions (of the same number, interval width, and composition of distal and genic/proximal regions as ACRs) were simulated using the BEDTools shuffle command to represent background noise, and the enrichment of ACR within each TE superfamily was assessed against the null distribution of control proportions based on permutation tests (n = 1000). Enrichment scores were calculated as the log2-transformed fold changes of observed versus the permutation-derived mean ACR proportions within TE superfamilies.

Differential accessibility (DA) analysis. Differences in chromatin accessibility attributable to hybridization and allopolyploidization were detected following an established DA workflow (Reske et al., 2020) using the R package csaw [v1.16.1] (Lun & Smyth, 2016). For direct comparison between different cotton species, all MNase-seq data were aligned to the same reference genome, either the AD1 reference genome or a concatenated reference of A2 and D5 genomes; DA results derived from both references were examined to mitigate bias. Mapped and quality filtered read pairs were counted into sliding windows or a given peak set to quantify MNase signals across the genome, followed by normalization based on the TMM or Loess method; multiple analytic approaches were evaluated to identify the most suitable DA workflow (Supplementary Text 4: Comparison of different analytic methods to conduct DA analysis). The resulting count matrices were then subject to the edgeR (M. D. Robinson et al., 2010) statistical framework of estimating dispersions by empirical Bayes and quasi-likelihood GLM fitting for hypothesis testing, according to the following designs: (a) light versus heavy in diploids; (b) light versus heavy in F1; (c) light versus heavy in AD1; (d) F1:light-heavy versus diploids:light-heavy, representing hybridization effect; and (e) AD1:light-heavy versus F1:light-heavy, representing polyploidization effect.

Motif discovery and enrichment analysis

Using the MEME Suite [v5.4.1] (Bailey et al., 2015) with default settings, scanning for known motif occurrences in the 1 kb promoter regions was conducted with FIMO (Grant et al., 2011), and combined motif discovery and enrichment analysis was performed using both XSTREME (Grant & Bailey, 2021) and AME (McLeay & Bailey, 2010). XSTREME conducts two types of de novo motif discovery using MEME and STREME followed by enrichment analysis using SEA (Bailey & Grant, 2021), and AME identifies known motifs that are relatively enriched in given sequences compared with control sequences. The promoter (<1 kb) ACRs per (sub)genome and corresponding promoter sequences were used as input and control sequences, respectively. The JASPAR core nonredundant plant motifs v2018 and Arabidopsis motifs from plantTFDB v5.0 (Jin et al., 2017) were used as known functional motifs. For clustering enriched motifs, the RSAT matrix-clustering tool (Castro-Mondragon et al., 2017) was used with the following parameters: -hclust_method average -calc sum -metric_build_tree Ncor -lth w 5 -lth cor 0.6 -lth Ncor 0.4 -quick. Heatmaps and hierarchical clustering were generated with Euclidean distance using the R package pheatmap (Kolde, 2017).

RNA-seq analysis

Total RNA extractions were performed using the Sigma spectrum plant total RNA kit (Cat No. STRN50), and quantified on a BioAnalyzer (Agilent, Palo Alto, CA). mRNA libraries were prepared using the Illumina

TruSeq RNA Library Prep Kit (Illumina, San Diego, CA, USA) and sequenced on three Hiseq 4000 lanes with paired-end 150-cycle sequencing. A total of 12 libraries from A2, D5, F1 and AD1 samples were generated with an average of 11 million read pairs per sample (Supplementary Table S1. Summary of sequencing data in this study). After quality filtering and trimming of adaptor sequences with TrimGalore (Krueger, 2012), paired-end reads were pseudo-aligned to the reference transcriptomes using Kallisto (Bray et al., 2016). Under the R environment version 3.5.0, differential gene expression analysis was conducted using DESeq2 (Love et al., 2014), with a false discovery rate $\alpha < 0.05$ required to identify significant changes.

To optimize the method to infer duplicated gene expression patterns, we tested the following mapping strategies. (1) D5-ref: The G. raimondii (D5) reference genome (Paterson et al., 2012) and a previously generated species-diagnostic SNP index (Page et al., 2013) were used to construct the reference transcript sequences for Kallisto mapping of RNA-seq data from each genotype. For this reference, D5 reads were mapped to the D5 transcripts; A2 reads were mapped to the “pseudo-A2” transcripts, which were generated by replacing species-diagnostic SNPs on the D5 gene models with A2-specific SNPs; F1 reads were mapped against a concatenation of the pseudo-A2 and D5 transcripts; and G. hirsutum (AD1) reads were mapped against a concatenation of pseudo-AD1-At and pseudo-AD1-Dt transcripts, which were similarly generated using AD1-specific SNPs. (2) AD1-ref: reads from all species were individually mapped against the G. hirsutum (AD1) transcript sequences (Chen et al., 2020a). (3) individual-ref: F1 reads were mapped to the concatenated A2 (Huang et al., 2020) and D5 (Paterson et al., 2012) transcripts, while A2, D5 and AD1 reads were each mapped to transcripts from their individual reference genomes. The resulting read counts from different references were compared based on syntenic ortholog/homoeolog relationships within the allopolyploid genome and between different references (i.e., A2, D5, F1:At, F1:Dt, AD1:At, and AD1:Dt), which were inferred using the pSONIC pipeline (Conover et al., 2021) as previously described (Conover & Wendel, 2022).

Based on the total (summed) expression of At and Dt homoeologs, F1 and AD1 gene expression was compared to expression in A2 and D5 and subsequently classified into following categories (Rapp et al., 2009): (1) additivity, whereby the total expression (in the hybrid or allopolyploid) is statistically equivalent to the mid-parent value of the parental diploids; (2) A-genome expression level dominance (ELD), whereby the total expression is statistically equivalent to the A2 parent but different from the D5 parent and mid-parent expression; (3) D-genome ELD, whereby the total expression is statistically equivalent to the D5 parent but different from the A2 parent and mid-parent expression; (4) transgressive up-regulation, whereby the total expression is greater than both A2 and D5; (5) transgressive down-regulation, whereby the total expression is less than both A2 and D5.

Based on the partitioned expression of At and Dt homoeologs (separately), homoeolog expression bias (HEB) was assessed in the F1 and AD1 by evaluating differential expression between homoeologs (At and Dt). Categorization of cis and trans regulatory divergence was performed as reported previously (Bao et al., 2019), which measured the overall contributions of cis and trans variants by log2 ratios of A2 and D5 (A = log2(A2/D5), the cis effects by log2 ratios of their corresponding homoeologs (B = log2(At/Dt)),

and then obtained the trans effects by A minus B. Based on the statistical significance of A, B, and A minus B, six categories of regulatory evolution were characterized as illustrated in Figure 6c. The evolutionary impact of hybridization (Hr), allopolyploidization (Pr), and genome doubling (Wr) were determined according to (Hu & Wendel, 2019) and as illustrated in Figure 6c.

Histone gene family analysis

Histone protein sequences of Arabidopsis thaliana were retrieved from HistoneDB 2.0 (Draizen et al., 2016) and Probst et al. (Probst et al., 2020), which were used as queries to search against cotton coding genes by BLASTP with e-5 as cutoff. Using the built-in functions of the Seaview version 5 software (Gouy et al., 2021), multiple sequences alignment was conducted using MUSCLE [v3.8.31] (Edgar, 2004), and phylogenetic analyses were performed using neighbor joining (NJ) and maximal likelihood (ML) methods. NJ trees were constructed with the “Poisson correction” model and a bootstrap test of 1000 replicates. ML trees were constructed using PhyML [v3.0] (Guindon et al., 2010) with the default “LG” model and 100 nonparametric bootstrap replicates. For each histone family, the average evolutionary divergence among family members was calculated in MEGA11 (Tamura et al., 2021) as the number of amino acid substitutions per site from averaging over all sequence pairs (i.e., overall mean distance) , using the Poisson correction model with all ambiguous positions removed for each sequence pair (pairwise deletion option).

Data and code availability

Data generated in this research are deposited in the NCBI short read archive: MNase-seq under PRJNA529909, ATAC-seq under PRJNA1018916, and RNA-seq under PRJNA529417. All data used are detailed in Supplementary Table 1. Custom scripts are available at the following GitHub repository: https://wendellab.github.io/cottonMNase-seq/.

Results

Mapping chromatin landscapes by differential sensitivity MNase-seq (DNS-seq)

To characterize the genome-wide chromatin features and cis-regulatory landscapes, we performed micrococcal nuclease (MNase) digestion of fixed chromatin in nuclei using two digestion conditions - heavy and light, titrated according to a previously established protocol (Vera et al., 2014). A total of 16 MNase-seq libraries were generated, consisting of two conditions for two biological replicates from four genotypes: the allopolyploid G. hirsutum cultivar Acala Maxxa (AD1; genome size (GS) = 2.2 Gb), A-genome diploid G. arboreum accession A2-101 (A2; GS = 1.8 Gb), D-genome diploid G. raimondii (D5; GS = 0.8 Gb), and their synthetic F1 hybrid (A2×D5; GS = 2.4 Gb)). An average of 60 million mono-nucleosome DNA-sized fragments (i.e. 150 bp read pairs) was sequenced per 1 Gb genome size per library, resulting in 591 million A2, 126 million D5, 549 million A2×D5, and 685 million AD1 read pairs (Table S2. Summary of MNase-seq dataset). After adapter trimming and quality filtering, the remaining 91-98% of reads were mapped to their corresponding reference genomes. Interestingly, the proportion of

high quality alignments (Q>20) was notably higher for the D5 reads (80-86%) versus the other genomes surveyed (range: 60-74%; Table S2. Summary of MNase-seq dataset), likely reflecting lower repetitive content of the smaller D5 genome. Quality evaluation of the mapping results indicates that genomic coverage profiles were highly correlated between biological replicates (R2=0.91~0.99); therefore, alignments from replicates (per species and per digestive condition) were combined in the following analyses.

As illustrated in Figure 1B (upper right), heavy digestion yields mainly mono-nucleosomes, as in traditional MNase-seq experiments, which enables genome-wide examination of nucleosome positioning and occupancy. The identification of well-positioned nucleosomes accounted for 16-20% of each Gossypium genome (Table S3. Nucleosome identification and classification), consistent with previous reports in human cells (Valouev et al., 2011) and plants (Wu et al., 2014; Zhang et al., 2015). The weakly positioned nucleosomes (or “fuzzy” nucleosomes), which accounts for 62-70% of each cotton genome, likely reflect positional variability and dynamics in multicellular samples. Around the transcription start site (TSS), the canonical pattern of nucleosome occupancy was observed: (1) the first nucleosome (+1 nucleosome) downstream of TSS is strongly localized, while array of phased nucleosome positioning gradually dissipates from the 5’ to the 3’ end of genes; (2) the region immediately upstream of the TSS is generally depleted of nucleosomes, and thus called the “nucleosome free region” (NFR), allowing access of TFs and other regulatory proteins; (3) highly expressed genes tend to have a lower degree of nucleosome occupancy and a larger NFR (Figure 1B).

The light MNase digestion releases more sensitive, “fragile” nucleosomes and sub-nucleosomal sized particles (e.g. transcription factors), which have been used to map MNase hypersensitive sites and profile chromatin accessibility as a complementary approach to DNase-seq and ATAC-seq (Parvathaneni et al., 2020; Pass et al., 2017; Savadel et al., 2021; Zhao et al., 2020). Here, the smaller DNA fragments (0-130 bp) sequenced from the light digestions were collected to identify open regions bound by subnucleosomal sized particles; while we refer to the corresponding genomic coverage from these light digestions as subnucleosomal particle occupancy (SPO; Figure 1B, lower left), as per (Teves & Henikoff, 2011), these regions are sometimes referred to as “MNase hypersensitive (MH) regions” (Zhao et al., 2020) or "MFs" for MOA-seq footprint (Savadel et al., 2021) regions. The Differential Nuclease Sensitivity (DNS; Figure 1B, lower right) approach permits identification of MNase sensitive footprints (MSFs) that reveal cis-regulatory landscapes (Parvathaneni et al., 2020; Rodgers-Melnick et al., 2016). Thus, we took a combined approach of examining genome-wide chromatin profiles including nucleosome occupancy each by light and heavy digestion, SPO, and MSF, for comparative analyses of ach Gossypium genotype studied (Figure S1).

To access chromatin accessibility profiles by MSF and SPO, we compared them with independent datasets generated using different enzymatic assays, ATAC-seq and DNase-seq (Table S4. Summary of ATAC-seq and DNase-seq data; Supplementary Text 2). Notably, the accessible chromatin regions (ACRs) detected by DNS-seq exhibited smaller peaks, lower GC content, and a more prominent distribution distal to genes compared to those detected by ATAC-seq and DNase-seq (Figure 1C-E; Table S5. ACR

identification in D5). Additionally, the accessibility profiles by MSF and SPO demonstrated the expected enrichment before TSS and depletion within the gene bodies, showing a positive correlation with gene expression levels (Figure S1), consistent with previous findings (Buenrostro et al., 2013). In contrast, ATAC-seq and DNase-seq did not exhibit such patterns (Figure S2). These results suggest that DNS-seq offers a good approach for mapping chromatin accessibility relative to gene expression levels in cotton. Furthermore, the distinct clustering by PCA (Figure S3) and limited overlaps in ACRs (Table S6. ACR overlap) between assays indicate that MSF and SPO profiles offer a unique view of genomic accessibility, particularly in distant non-genic regions which appear to be less well represented in assays based on Tn5 or DNase I (Figure 1E). This observation is consistent with the previous findings in Arabidopsis (Pass et al., 2017; Zhao et al., 2020).

Alteration of nucleosome organization by hybridization and allopolyploidization

To compare nucleosome organization between diploid, hybrid, and allopolyploid cottons, we first computed phasograms to analyze the global patterns of nucleosome positioning and spacing. A phasogram represents the frequency distributions of distances between mononucleosomal reads mapped (i.e. from heavy MNase digestion), observed as oscillating sine wave signals, for which period is the center-to-center distance between neighboring nucleosomes, averaged genome wide (Valouev et al., 2011). For each cotton genome, the average distance between neighboring nucleosomes, also known as nucleosome repeat length (NRL), was estimated by applying a linear model to calculate the phasogram period (Figure 2A; Figure S4; Table S7. NRL estimation). Interspecific and intergenomic comparisons revealed subtle but statistically significant genotype-based variation in average nucleosome spacing. We found that NRLs were generally shorter in the diploids and the diploid-hybrid (F1) versus the allopolyploid cotton (AD1) and that the D-genome NRLs were generally shorter than those in the A-genome (Figure 2B; Diploids: A2 197.3 ± 0.2 bp, D5 196.2 ± 0.5 bp; F1: At 197.5 ± 0.2 bp, Dt 196.4 ± 0.4 bp; AD1: At 200.1 ± 0.4 bp, Dt 199.7 ± 0.5 bp; ANOVA followed by Tukey's post hoc test, $P < 0.05$: diploids= F1< AD1 and D<A). Consistent with these observations, the percentage of genomic regions occupied by nucleosomes (i.e., nucleosome coverage, NC) also exhibited lower A- versus D- coverage, regardless of ploidy (Figure 2C; Diploids: A2 78.2 ± 0.5 %, D5 89.9 ± 0.1 %; F1: At 78.8 ± 0.6 %, Dt 90.0 ± 0.21%; AD1: At 80.1 ± 0.3 %, Dt 84.1 ± 0.4 %; A<D, Student's T test $P < 0.05$). These results, i.e., shorter NRL and higher NC in the D-genome, together indicate that nucleosomes are generally arranged further apart in the larger A genome. Furthermore, both the NRL and NC reveal significantly larger interspecific differences between the A2 and D5 diploids relative to the inter-subgenomic differences between At and Dt in the allopolyploid (AD1), which suggests that allopolyploidization and subsequent evolution as a tetraploid, but not hybridization per se, may result in homogenization of nucleosome density.

Nucleosome positioning is known to be directed by a combination of the intrinsic properties of DNA sequence that act in cis and chromatin remodeling that deploys transcription machinery that acts in trans (Radman-Livaja & Rando, 2010). Therefore, we next examined the roles for cis- and trans-acting factors in changing the nucleosome distribution during genome evolution. To isolate the cis effects, we applied a sequence-based computational model to predict the "intrinsically DNA-encoded" nucleosome features. If

each prediction agrees with the experimental estimation, we conclude that cis DNA sequence plays a significant role; otherwise, a significant trans effect would be inferred. Interestingly, sequence-based predictions of nucleosomal spacing and coverage for each reference genome (i.e., A2, D5, and AD1) suggests that the NRL should be longer in the A-(sub)genomes (versus the D-(sub)genomes) with a concomitantly lower NC value, regardless of ploidy level (Figure 2D and 2E). This observation directly contrasts the experimentally observed pattern (Figure 2B and 2C) and therefore implies a possible role for trans effects in nucleosome positioning. Given this observation, it is perhaps surprising that the sequence-based nucleosome positioning predictions for diploid versus polyploid cotton mirrored that of the MNase-seq estimations, both of which find that the differences in NRL and NC between the A2 and D5 diploids exhibit significant reductions in the At- and Dt-subgenomes of the allopolyploid (AD1). In other words, the synchronization effect on nucleosome organization was impacted in cis by sequence evolution accompanying allopolyploidization.

Chromatin accessibility increases in allopolyploid promoters

ACRs were identified for each sample from the DNS and SPO data combined, comprising 1.1-1.4% of each genome (Table 1; Table S8-S11. ACR identification per chr). In the F1 hybrid, we identified 581,654 ACRs covering 30.9 Mbp. Both the numbers and total genomic fractions of ACRs were higher than their combined counterparts in the diploid progenitors, A2 (296,312; 16.4 Mbp) and D5 (190,795; 9.2 Mbp). In the allopolyploid AD1, only the total length of ACRs (449,346; 27.4 Mbp) surpassed that of diploid progenitors. A majority of ACRs were located >2 kb from their nearest gene (distal, dACRs: 72-87%), whereas 10-19% occurred proximally within the 2 kb gene flanking regions (proximal, pACRs) and only 4-12% overlapped gene bodies (genic, gACRs). The larger A-(sub)genomes exhibited a higher proportion of dACRs and commensurately lower proportions of gACRs and pACRs relative to the smaller D-(sub)genomes (Figure 3A), consistent with observations in other plant species which suggest that the proportion of dACRs is positively correlated with genome size (Lu et al., 2019). This correlation with genome size was even more significant for the total length of dACRs (Figure 3B), whereas gACRs and pACRs were mostly comparable between A- and D-(sub)genomes, likely due to their general conservation in genes. Interestingly, the proportion and total length of pACRs was significantly increased in AD1, specifically due to expansions in the 1 kb promoter regions (Figure 3C and 3D).

Within 1 kb promoter regions, an initial scan of each genome for known DNA motifs from plantTFDB v5.0 (Jin et al., 2017) revealed relatively consistent motif occurrences across (sub)genomes, although the A2 promoters exhibited the most divergence relative to the other genomes (Figure 4A), possibly due to the elevated GC content in its promoters (A2 30.56%, versus AD1:At 28.08%, D5 28.72%, AD1:Dt 28.98%). We then used this background variation in 1 kb promoter sequences as a control to obtain enriched motifs from the pACRs by AME, resulting in 351, 326, and 408 enriched motifs in the parental diploids (aggregated), the F1, and in AD1, respectively (Supplementary Table 12. Ranking of significantly enriched motifs in 1 kb promoter ACRs). Among the union of 423 enriched motifs, 247 were shared by all (sub)genomes, indicating a high level of cis-element conservation among cotton (sub)genomes (Figure 4B). Interestingly, AD1-specific motifs comprised the second, fourth, and fifth largest intersecting sets,

which include 33 motifs enriched in both At and Dt pACRs, 15 enriched in Dt only, and 13 enriched in At only. These motifs mostly belong to TF binding sites of MYB (10 motifs), WRKY (9 motifs), bZIP (9 motifs), and TCP (9 motifs) transcription factor families (Figure 4C; Supplementary Table 12). Congruently, a heatmap dendrogram of pACR motif enrichment rankings showed that AD1:At and AD1:Dt were more similar to each other and distinct from the diploid enrichment rankings. Among the diploids, clustering of F1:Dt and D5 showed their higher similarity, with the A2 and F1:At genomes falling more basally in that clade (Figure 4B). Furthermore, de novo motif discovery by XSTREME and clustering analysis (Figure S5) confirmed these patterns, suggesting a synchronization effect associated with allopolyploidization and a potentially asymmetric effect associated with hybridization.

Decreased chromatin accessibility in repetitive regions accompanying allopolyploidy

Genome-wide characterization of transposable elements (TEs) revealed that the A subgenome of AD1 has 1.2% lower TEs than A2 whereas the D subgenome has 5.1% more TEs than D5 (AD1:At = 81.2%, AD1:Dt = 64.6%, A2 = 82.4%, and D5 = 59.5%; (Figure 5A; Supplementary Table S13), consistent with previous reports (Chen et al., 2020b; Zhao et al., 1998). ACRs accounted for only 0.31-0.68% of genomic regions annotated as TEs, significantly lower than their composition in other genomic regions (1.09-1.37%; permutation test $P < 0.05$). Depletion of ACRs was evident for all TE superfamilies, with the greatest depletion detected for the Gypsy retrotransposons (Figure 3E), as expected by their general tendency to reside in heterochromatic regions. More A- than D- (sub)genomic ACRs overlapped with TEs, particularly LTR retrotransposons, congruent with the higher TE content in the larger A-genome (Table 1). Regardless of subgenome, however, the allopolyploid (AD1) contained the lowest amounts of ACRs that overlapped with TEs (AD1: At 22.5%, Dt 17.6%; F1: At 35.8%, Dt 26.8%; A2 32.5%; D5 22.4%), indicating decreased chromatin accessibility in TE regions accompanying allopolyploidization.

Because the allopolyploid (AD1) exhibits both a reduction in TE-overlapping ACRs and an increase in promoter ACRs (Figure 3D), we hypothesized that promoters may have gained more accessibility from TE removal associated with polyploidization. The general distribution of TEs around transcription start sites (TSS) is similar between diploid and polyploid cottons (Figure S6). However, the diploid A2 exhibits a strikingly high number of Gypsy elements within its 1 kb promoter regions that are absent from its homologous genome in the allopolyploid, and this pattern was not observed in other genomic regions (Figure 5B; Figure S7; Table S14). At the genome-wide scale, TEs contributed to 18-36% of ACRs (Table 1; Figure 5C), but these accessible TEs were mainly located in distal intergenic regions and only contributed to a small portion of promoter ACRs (Figure 5D). The DNA transposon Mutator-derived ACRs were most abundant within promoters, consistent with their genomic distribution and tendency to be near genes compared to the distribution pattern of¬ LTR retrotransposons. Interestingly, the loss of Gypsy in AD1 promoters is associated with a gain of both non-TE and TE-derived ACRs accompanying allopolyploidy (Figure 5A-D). Although this observation supports our hypothesis that promoter TE depletion led to increased accessibility in the At genome of the allopolyploid (relative to A2), it does not explain the increased accessibility in the Dt genome (versus D5).

Because TE superfamily distribution may vary among genomic regions, we asked whether any particular TE families represented a key source of ACRs. Although we observed a strong positive correlation between genome-wide TEs and TE-derived ACRs for superfamilies within each genome (AD1:At 0.96, AD1:Dt 0.84; F1:At 0.98, F1:Dt 0.96; A2 0.99; and D5 0.93), we did not observe ACR enrichment of particular TE superfamilies. Out of the 28,057 TE families characterized across cotton species, a union of 8,680 families were found significantly enriched in TE-derived ACRs (Figure 5E). Intersection of TE families among genomes revealed a significant proportion of lineage-specific TE families, which accounts for 14-20% of the TE -overlapping ACRs in each (sub)genome. The largest intersection set of 897 families was only found in the At genome of the diploid synthetic hybrid. While these lineage-specific families are mainly LTR retrotransposons, TE families shared by at least half of the genomes tend to be depleted of Gypsy and enriched in Mutator and hAT.

Because TEs are often associated with both inaccessible chromatin and transcriptional repression, we evaluated the expression of accessible TEs using transcriptomic data. We found that TE-based transcripts from 7,045 TE families accounted for 3-6% of mapped RNA-seq reads. Notably, 4622 of these expressed TE families also significantly contributed to ACRs (i.e., found among 8,680 families mentioned above). The significant overlap between transcription and accessibility (chi-square association test $P < 0.05$) indicates that accessible TEs are likely to be transcriptionally expressed. Interestingly, while the numbers of expressed TE families were comparable between cotton genomes (A2 3622, D5 2973, F1:At 3242, F1:Dt 3115, AD1:At 3381, AD1:Dt 2946), higher transcript abundances were found in D5 (Figure 5F; Figure S8).

Allopolyploidy causes more accessibility changes than does hybridization

Both interspecific hybridization and polyploidization can have profound effects on the epigenome and gene expression. To assess their effects on chromatin accessibility, we initially compared ACRs identified in the diploid A2 and D5 genomes with their homologs in the At and Dt subgenomes of the interspecific diploid hybrid (F1). Surprisingly, this comparison resulted in little overlap between diploid and F1 ACRs. While this could indicate substantial changes in genome-wide chromatin accessibility due to hybridization, it more likely reflects technical issues in comparing independently identified ACRs with high stringency across genomes.

To address this issue, we employed a differential accessibility (DA) analysis directly contrasting the MNase-seq data between diploid genomes and their corresponding subgenomes in the F1 and natural allopolyploid (see Methods and Supplementary Text 4: Comparison of different analytic methods to conduct DA analysis). The DA analysis of allopolyploid versus diploids revealed an increase of 3.3-4.4 Mb and a decrease of 435-740 kb in accessibility. In contrast, the differences between the F1 hybrid and the diploids were smaller, showing an increase of 16-304 kb and a decrease of 132-214 kb. These results indicate that allopolyploidization and subsequent evolution at the allopolyploid level, for > 1-2 million years in this case, collectively induce much greater changes in chromatin accessibility than does

hybridization in the F1. The consequences of these accessibility changes, particularly in promoter regions, were explored next.

Duplicated gene expression in diploid hybrid and allopolyploid cotton

To assess the consequences of chromatin changes on gene expression evolution, we first characterized the evolution of duplicated gene expression using matching RNA-seq data generated for the two diploids A2, D5, their F1 hybrid, and natural allopolyploid derivative, AD1 (Supplementary Table S1. Summary of sequencing data). Duplicated gene expression patterns were categorized under a pre-established analytical framework (Hu & Wendel, 2019), illustrated in Figure 6. We employed a conservative approach and only report results that are consistent across different mapping strategies (Supplementary Table S15. Mapping summary of RNA-seq data; Supplementary Text 5. Analysis of duplicated gene expression patterns using different mapping strategies; Supplementary Table S16-18. Summary table of duplicated gene expression patterns). We also restricted our analysis to genes where orthology and homoeology among (sub)genomes could be confidently determined. For each of these 22,889 ortho-homoeolog groups (OGs; each containing a single representative for A2, D5, F1:At, F1:Dt, AD1:At, and AD1:Dt), duplicated gene expression patterns were characterized based on total (Figure 6A-B) and partitioned homoeologous expression levels (Figure 6C-D), including differential total expression relative to parental diploids, expression level dominance (ELD), homoeolog expression bias (HEB), cis and trans regulatory divergence, as well as the evolutionary impact of hybridization (Hr), allopolyploidization (Pr), and genome doubling (Wr).

In both the F1 and AD1, the total expression of homoeologous genes exhibited more differential expression relative to A2 than to D5 (F1 - 13.7% versus 8.0%; AD1 - 11.5% versus 7.0%; Figure 6A), and, correspondingly, the ELD analysis revealed more D-dominant than A-dominant expression patterns (F1 - 5.8% versus 2.3%; AD1 - 4.4% versus 2.2%; Figure 6B). These observations suggest an asymmetric resemblance of the overall transcriptome towards the D-genome diploid parent, as noted previously (Flagel et al., 2008; Rapp et al., 2009; Yoo & Wendel, 2014). This trend was consistent across different mapping strategies (Table S16-18).

When expression was compared between homoeologs, HEB was detected for 6.5% ($B \neq 0$) and 14.1% ($Bp \neq 0$) of genes in the F1 and AD1, respectively (Figure 6D; Table S18), representing a greater than twofold increase in HEB in the allopolyploid. While no overall imbalance in HEB was detected in the F1, in the allopolyploid more homoeologous pairs exhibited D- (versus A-) biases, regardless of mapping strategy. Allele-specific expression (ASE) analysis revealed 12.8% of genes exhibited parental expression divergence between A2 and D5 , whose inferred regulation can be subdivided into the previously described categories (Hu & Wendel, 2019): cis only (I - 849 genes), trans only (II - 62 genes), cis and trans enhancing (III - 9 genes), and cis and trans compensating (IV - 8 genes). Notably, these results ascribe an order of magnitude greater influence of cis variation in expression evolution between the diploid cottons (Table S18), suggesting that cis evolution has played a dominant role in generating expression variation

between those species. In terms of the evolutionary impact of genome polyploidy, genome doubling (Wr ≠ 0, 2.9%) has a much stronger effect than hybridization (Hr ≠ 0, 0.6%), representing two distinct phases of allopolyploidization (Pr ≠ 0, 4.7%). These results also suggested that the relative expression of At versus Dt homoeologs in F1 and AD1 was mainly determined by the parental state of A2 versus D5 (Hr = 0, 99.4%; Pr = 0, 95.3%), also known as “parental legacy” (Buggs et al., 2014). Only a small portion of At versus Dt ratios were distinct from the parental states, a situation known as “regulatory novelty” by hybridization and allopolyploidization (Hr ≠ 0 and Pr ≠ 0, as illustrated in Figure 6C).

Promoter accessibility regulates duplicated gene expression patterns

To explore the links between chromatin architecture and expression evolution, we next examined the promoter accessibility profiles as measured by DNS signals in association with various duplicated gene expression patterns. For a total of 22,889 orthogroups (OGs, see above), direct comparison of the A2 versus D5 parental profiles revealed a systematic shift around transcription start sites, likely due to differences in gene annotation between the two diploid reference genomes (Figure 7A, top row). This technical issue limits our ability to directly detect accessibility changes associated with parental expression divergence (A ≠ 0). In contrast, the use of the allopolyploid reference genome revealed that promoter accessibility is positively correlated with homoeologous expression levels; that is, higher A- versus D- promoter accessibility was observed for the homoeologous gene pairs exhibiting A-biased HEB (Bp ≠ 0), and higher D- versus A- promoter accessibility was observed for pairs exhibiting D-biases (Figure 7A, bottom row). Additionally, the homoeolog that exhibited biased higher expression tended to display larger ACRs within 1 kb of the TSS (Figure 7B). For HEB in F1 (B ≠ 0), interestingly, the use of diploid (A2 and D5 concatenated) and allopolyploid (AD1) references both revealed a systematically higher A- than D- promoter accessibility, regardless of the direction of HEB (Figure 5A, middle two rows). Although we cannot rule out artifacts introduced by either reference, the distinct patterns in diploid hybrid vs. allopolyploid cotton indicate that hybridization alone does not alter the relationship between gene expression and promoter accessibility, but the allopolyploid evolution does. In addition to OGs, we also characterized promoter accessibility for genes that cannot be confidently assigned to ortho-homoeolog groups (referred to as nonOGs: A2 - 18850; D5 - 14329; At - 13227; Dt - 15895) and found distinct patterns between OG and nonOG genes (Figure S9). In allopolyploid cotton, a higher A- than D- accessibility was shown for all genes and nonOGs, whereas comparable A- and D- accessibility levels were shown for OGs.

Given the strong evidence of “parental legacy” of hybridization with respect to both nucleosome organization (Figure 2) and gene expression (Figure 6), we hypothesized that in the hybrid, promoter accessibility and its regulatory consequences on gene expression would be primarily vertically inherited and thus mirror parental profiles. To test this hypothesis, we examined the relative A- versus D- genome chromatin accessibility profiles for categorized expression patterns, by normalizing the A-genome profiles (A2, F1:At, and AD1:At) and D-genome profiles (D5, F1:Dt, and AD1:Dt) against their corresponding diploid references by genomic content (Figure S10 and Figure S11). The results revealed a relatively slight decrease by hybridization and a much stronger increase by allopolyploidization in accessibility (F1 < A2/D5 << AD1) for both the A- and D- genomes, as evident in aggregation plots (Figure 7D) and DA tests

(Figure 7C). The DA results also indicated more accessibility increases in At versus Dt promoters, consistent with the previous marginal comparison of ACRs (Figure 3C and 3D). In association with the impact of genome doubling on gene expression, the up-regulation of At/Dt homoeolog expression ratios (Wr > 0) was attributed to a biased increase in At promoter accessibility, while the down-regulation of At/Dt homoeolog expression ratios (Wr < 0) were attributed to a biased increase of Dt promoter accessibility (Figure 7E). No apparent accessibility patterns were observed with the impact of hybridization (Hr ≠ 0 in 142 OG; Figure S10 and Figure S11A), likely due to small changes. These results show that “parental legacy” can be seen with chromatin structural features, implicating cis-regulation as a heritable feature of promoters in different genotypic backgrounds.

With respect to non-additive patterns accompanying hybridization and polyploidy, we investigated how promoter accessibility changes of At and Dt homoeologs were associated with ELD in F1 (Figure S11B) and AD1 (Figure S11C); this analysis is summarized in Figure 7F. In the diploid hybrid, when higher parental expression was detected in A2 than D5, hybridization appeared to further increase the promoter accessibility of At to establish the A-dominant ELD, and decrease the promoter accessibility of both At and Dt to establish the D-dominant ELD. Conversely, when higher parental expression was detected in D5 compared to A2, the hybridization appeared to further increase the promoter accessibility of Dt to establish the D-dominant ELD. The same was true for A-dominant ELD. Therefore, the regulatory effect of chromatin accessibility changes primarily affects the homoeolog with higher parental expression in F1. Interestingly, in allopolyploid cotton, accessibility changes were primarily in At promoters, likely due to sequence evolution accompanying natural allopolyploidization (Figure 7F). These results demonstrate the distinct regulatory evolution accompanying hybridization versus allopolyploidization.

Histone gene expression evolution in association with nucleosome organization as mediated by chromatin accessibility

Because histone proteins are essential for nucleosome assembly, we next focused on histone gene expression to ask whether their expression levels vary between (sub)genomes and across ploidy levels, and how this relates to the observed nucleosome spacing patterns. In G. hirsutum, we identified 149 histone coding genes, including variants of core histones (H2A - 24 At and 23 Dt, H2B - 13&13, H3 - 18&18, and H4 - 14&16) and linker histones (H1 - 5 At and 5 Dt), based on phylogenetic relationships and amino acid sequence similarities with 50 well-characterized histone genes in Arabidopsis (Supplementary Table 19). Estimates of the average evolutionary divergence for each family revealed that H1 and H2A comprise more divergent variants than the other families (overall mean amino acid distance: H1 - 0.53, H2A - 0.44, H2B - 0.16, H3 - 0.08, and H4 - 0.02), consistent with previous findings in animals and plants (Probst et al., 2020).

To investigate the expression patterns of histone genes, we examined 47 OGs containing genes from the A2, D5, AD1:At and AD1:Dt genomes (H1 - 4, H2A - 17, H2B - 7, H3 - 8, and H4 - 11) (Supplementary Table 20). We found that the total expression of these genes were higher in the allopolyploid (mean sum TPM with standard deviation: 4988.4±189.5) compared to the diploid genomes (3714.3±301.1 in F1,

4338.3±414.9 in D5, and 3207.2±661.2 in A2), which agrees with the expectation that the allopolyploid genome contains more nucleosomes than do diploid genomes, such that histone transcription needs might be greater on a per cell basis. At the histone gene family level, notably, this increase in expression was particularly evident for linker histone H1 (Figure 8A). At the histone variant level (Figure S12), this pattern was observed for canonical H1, H2A.X, H2A.Z variants , due to transgressive up-regulation of their gene members in allopolyploid cotton; interestingly, their counterparts in hybrid F1 tend to exhibit the D5-like expression .

Analyzing the expression difference between A- and D- (sub)genomes, we generally observed higher expression in the D-(sub)genome despite the overall lack of statistical significance (Figure 8A; Figure S12). Assuming histone expression levels correlate with the nucleosome number, it is intriguing that the smaller D5 diploid exhibited statistically equal, or even higher histone expression levels compared to the much larger A2 diploid (genome size 0.8 Gb vs 1.6 Gb). This finding is consistent with the nucleosome positioning result, i.e., smaller D5 diploid exhibiting higher nucleosome coverage, suggesting a speculation that histone gene expression may contribute to regulation of nucleosome spacing.

At the OG level (i.e. expression by gene), more non-additive expressions were detected in the allopolyploid than hybrid (Figure 8B). Directions of parental divergence (i.e. A > 0 and A < 0) and HEB (i.e. B > 0 and B < 0, or Bp > 0 and Bp < 0) were more or less balanced, which were often influenced by cis-only regulation, and no significant trans regulatory divergence was detected (Figure 8C). For instance, a larger and more prominent promoter ACR region was found associated with higher expression of the canonical H1 gene in the At subgenome compared to the Dt subgenome (Figure 8D-E: OG0025113 - Gohir.A13G169300 versus Gohir.D13G174801). More examples were shown in Figure S13.

Discussion

In this study we employed the MNase-based DNS-seq technique to examine chromatin structural features in the context of allopolyploid cotton, G. hirsutum, to address two primary questions regarding the evolutionary impact of allopolyploidization: (1) how does genome merger and doubling accompanying allopolyploidy alter chromatin structure; and (2) what evidence can be obtained that connects the regulatory aspects of chromatin structure to the evolution of duplicated gene expression?

Dissecting cis and trans determinants of polyploid chromatin evolution

With respect to the first question, our data suggest stronger effects on the genome-wide chromatin landscape by allopolyploidy than by hybridization (Pr ≫ Hr) , noting that the former entails both genome doubling and, in the case of Gossypium, 1-2 million years of natural evolution as the lineage diversified and spread across the many regions in the American tropics. Notably, a preponderance of chromatin alterations appear to have been driven by sequence evolution acting in cis. First, relative to the parental A2 and D5 diploids that model the allopolyploid progenitors, only slight changes of nucleosome organization (Figure 2B-C) and chromatin accessibility (Figure 7C) were detected in the F1 hybrid, with allele-specific patterns closely mirroring those of diploids. This lack of deviation from vertical

transmission of pre-existing chromatin patterns clearly indicates strong “parental legacy” (Buggs et al., 2014) by hybridization, as well as the cis nature of parental divergence on chromatin features, in accordance with the classic allele-specific expression (ASE) model (Wittkopp et al., 2004).

Next, a multilevel synchronization effect was evident in the allopolyploid, which has assimilated various sequence-based and chromatin level features of both the A and D progenitor genomes, including nucleosome spacing (Figure 2B-C), ACR classification (Table 1), genomic TE content and distribution (Figure 5A and Figure S6), accessibility round TSS (Figure 7A), and the promoter cis-regulatory landscape (Figure 4). These results are consistent with the previous study of genome-wide chromatin analysis in diploid and polyploid cottons using DNase-seq and further enrich the evidence of synchronization effects based on DHS accessibility and histone modification marks (Han et al., 2022).

Notably, although the synchronization effect accompanying allopolyploidy resembles the trans effect in the synthetic hybrid, it cannot be simply interpreted according to the classic ASE model. As previously proposed (Hu & Wendel, 2019), an extended cis-trans framework is required to delineate the cis and trans determinants of gene expression that arise from genome doubling following hybridization. That is, under the common trans environment experienced by both subgenomes in the allopolyploid, the partitioning of cis-trans regulation needs to be conceptually modeled into inter- and intra- subgenomic interactions, based on integrated analysis of genetic and epigenetic variations. While more sophisticated computational modeling and molecular tools are needed to fully elucidate these interactions, we demonstrated the use of computational prediction to pinpoint cis determination of nucleosome positioning (Figure 2D-E), where reduced difference in nucleosome spacing by allopolyploidy can be predicted by DNA sequence per se. It has been recognized that nucleosome formation favors periodic distribution of the dinucleotides GG, TA, TG and TT at contact points between DNA and histones (every ~10 bp) and sequences such as poly(dA:dT) that require high DNA bending energy tend to be avoided (Kaplan et al., 2009; Segal & Widom, 2009). Therefore, nucleosome positions represent sequence-encoded functional features, which can therefore be selected during evolution (Barbier et al., 2021). We hypothesize that subgenomes in allopolyploids could be differentially selected (toward convergence) not only for their homoeologous gene content, but also for their ability to favor or impair nucleosome formation at genome-wide scale to facilitate chromatin package and/or at specific loci to impact accessibility to regulatory factors that mediate selectively favored gene expression. Future studies involving additional allopolyploid systems and tissue types will be instrumental in this hypothesis of nucleosome evolution.

In contrast to the cis determination of synchronization in terms of nucleosome spacing and promoter accessibility, the characteristics of nucleosome positions turned out to be strongly shaped by trans factors, as evidenced by disparity between experimental observations and DNA predictions (Figure 2). That is, distances between consecutive nucleosomes were greater in A- than in D- (sub)genomes, whereas the opposite patterns were suggested by the computational prediction of nucleosome occupancy from DNA sequences alone. With a fixed length of ~147 bp for canonical nucleosomes, NRL ranges from 154 bp in fission yeast (Lantermann et al., 2010) to 240 bp in echinoderm sperm (Athey et al., 1990),

depending on species, tissue type, and experimental conditions. Studies on yeast, animal, and human have shown that NRL tends to be shorter in transcriptionally active genomes, such as embryonic stem cells and tumor cells compared to echinoderm sperm, or active gene regions compared to heterochromatic non-coding sequences (Barbier et al., 2021). Notably, telomeric chromatin stands as an exception to this rule, exhibiting an unusually short NRL and high sensitivity to MNase (Tommerup et al., 1994) due to its unique columnar conformation of nucleosome stacking (Soman et al., 2022).

In plants, MNase digestion analysis of cereal species has revealed a typical NRL of 175-185 bp, with shorter NRLs observed in telomeric nucleosomes compared to bulk nucleosomes (Vershinin & Heslop-Harrison, 1998). Additionally, intriguing differences in MNase kinetics were observed between rye (7.8 Gb, 2n=14) and wheat (160 Gb, 2n=42), where the shorter NRL and faster MNase cleavage of the smaller rye genomes was proposed to be influenced by its prominent subtelomeric heterochromatin. Recent phasogram analyses using mononucleosomal MNase-seq have also been conducted in Arabidopsis (135 Mb; NRL of 185.1 bp in leaves and 182.2 bp in flowers), rice (430 Mb; 188 bp in leaves), and maize (2.4 Gb;193.5 bp in shoots and 190.7 bp in endosperm) (Chen et al., 2017; Zhang et al., 2015), further supporting the trend of larger nucleosome spacing in larger genomes, as observed here for cotton. In both rice and arabidopsis, heterochromatic regions were found to have larger nucleosome spacing compared to euchromatic regions marked by various histone modifications (Zhang et al., 2015). Similarly, in maize, intergenic regions exhibited larger spacing than the genome-wide NRLs (Chen et al., 2017). Differential spacing of nucleosomes associated with distinct genomic regions has also been reported in the human genome (Valouev et al., 2011). Such variations of NRLs have been well recognized to direct the folding of nucleosome arrays into chromatin fibers (Brouwer et al., 2021; Fransz & de Jong, 2011): evidently, longer linker DNA (197 bp vs 167 bp) together with the binding of linker histones (H1, H5) are required for a further compaction and stabilization of the 30 nm chromatin fiber, as associated with a repressed chromatin state. Indeed, we identified significantly higher expression levels of the linker histone H1 corresponding to larger NRLs in A- versus D- (sub)genomes, as well as the allopolyploid versus diploids.

Hence, it is plausible that plant genomes with larger sizes and higher ploidy levels have undergone adaptations resulting in larger nucleosome spacing, potentially facilitating specific high-order chromatin organizations. Additional studies are necessary to test this hypothesis. Apart from the cis-regulatory role of DNA sequences in nucleosome organization, there are several trans factors that contribute to this process, including histone variants, post-translational histone modifications, chromatin remodeling enzymes, and various architectural proteins (Arya et al., 2010). To fully understand the complex interplay between cis and trans elements in shaping nucleosome organization in polyploid plant genomes, it will be crucial to investigate the sequence and functional evolution of these factors accompanying allopolyploidization.

Regulatory relationships among chromatin evolution and duplicated gene expression

To address our second main question, above, regarding regulatory control of gene expression evolution accompanying allopolyploidization, we investigated the role of promoter accessibility in shaping various

well-recognized phenomena of duplicated gene expression, including asymmetric resemblance of parental diploids, homoeolog expression bias (HEB), nonadditive inheritance modes, and genome impact of hybridization (Hr) and allopolyploidization (Pr). Central to this investigation was also the extended cis-trans analytic framework (Hu & Wendel, 2019), which enabled us to first systematically characterize these duplicated gene expression patterns (Figure 6), and next disentangle the regulatory effects of chromatin accessibility (Figure 8). By exploring interconnecting patterns among chromatin traits and duplicate gene expression patterns, our study provides several perspectives into the regulatory underpinnings that govern allopolyploid gene expression dynamics.

Regulatory Relationships to Homoeolog Expression Bias. The positive correlation between promoter accessibility and gene expression levels reaffirmed the anticipated connection between HEB direction and accessibility in the allopolyploid; that is, the homoeolog exhibiting higher expression level exhibits greater promoter accessibility than its alternative duplicated copy. However, this regulatory connection was not observed in the synthetic diploid hybrid, which exhibited a systematic asymmetry of higher A- than D-promoter accessibility, irrespective of HEB direction (Figure 7A). This observation suggests that hybridization by itself generates “mismatches” between gene expression and chromatin accessibility, raising intriguing questions about the temporal scale and mechanisms in establishing their regulatory relationships during allopolyploid formation and evolution. One other implication is that HEB is determined by chromatin features or transcriptional factors other than or in addition to promoter accessibility.

The Temporal Scale of Regulatory Evolution. Assessment of Hr and Pr revealed contrasting effects of immediate hybridization and evolution of the cognate allopolyploid lineage. Hybridization is shown to be characterized primarily by parental legacy, manifested as mostly “vertical inheritance” of expression levels with minor changes in both accessibility and expression. In contrast, allopolyploidization exerts a pronounced impact, leading to substantial accessibility increases attributed to genome doubling and subsequent sequence evolution. Furthermore, the homoeolog-specific accessibility increase was notably associated with shifts in homoeolog expression ratios (e.g., Wr >0 or Wr < 0 in Figure 7E), underlining the regulatory influence of chromatin dynamics. Our promoter analysis highlights the potential role of sequence evolution in reducing TE contents and introducing cis-regulatory footprints into gene promoter regions, thereby impacting chromatin accessibility and gene expression evolution. Relationships between these dynamics and the multiple cascading spatial and stoichiometric effects of genome doubling (Bottani et al., 2018; Doyle & Coate, 2019) comprise a promising direction of future research.

Non-additive Inheritance Modes. Although allopolyploidization led to accessibility increases, we did not detect a significant amount of transgressive up-regulation of gene expression relative to parental diploids, as might have been expected. This observation implicates additional regulatory influences and perhaps stoichiometric controls on gene expression, the identification of which also comprise an interesting research direction. The phenomenon of expression level dominance (ELD), another well-known yet mechanistically mysterious non-additive expression pattern, perhaps exemplifies the complexities of the interplay between chromatin accessibility and gene expression. Our study demonstrates that changes in

chromatin accessibility predominantly impact the homoeolog with higher parental expression in the F1 generation; in contrast, allopolyploidy is characterized by a distinctive pattern in which accessibility changes predominantly occur in At promoters, a shift likely driven by sequence evolution linked to natural allopolyploidization (Figure 7F). Yoo et al. (Yoo et al., 2013) previously investigated homoeolog expression levels relative to ELD patterns and also showed that ELD reflects the up- or downregulation of alternative homoeologs more frequently, compared to the up- or downregulation of both homoeologs. The interrelationships among these dynamics remain to be elucidated.

Concluding Remarks

Here we show that promoter accessibility and nucleosome arrangement represent key components of the evolution of duplicate gene expression. It is important to acknowledge, though, that the realm of “chromatin structure” encompasses multiple molecular biological, quantitative, and spatial dimensions, with numerous mechanisms yet to be integrated into the needed synthesis. For instance, the interplay between DNA methylation and chromatin accessibility remains to be further elucidated in response to hybridization and polyploidization. Between the parental diploids, the D-genome G. raimondii contains more TEs near genes than does the A-genome G. arboreum, and hence G. raimondii orthologs were generally more methylated (Song et al., 2017). Upon hybridization, CG and CHG methylation levels were conserved whereas CHH methylation levels were decreased in the synthetic F1, and the majority of these changes were conserved during the subsequent polyploid evolution. In the allopolyploid cotton, however, more CG methylation and lower euchromatic H3K4me4 levels (Zheng et al., 2016) were found in the At than Dt homoeologs, in association with more D-biased HEB. While our work also detected a significant imbalance of D-bias in AD1 (Figure 6D), the globally higher promoter accessibility in the A- than D-genome remains enigmatic.

The orchestration of three-dimensional chromatin organization is another crucial facet of chromatin evolution. Alterations in spatial subgenome distribution into different genome territories and long-range interactions within and between subgenomes intricately link to homoeologous gene expression (Pei et al., 2021). In cotton, allopolyploidization led to chromatin compartment switching and topologically associated domain (TAD) reorganization, both influencing gene expression dynamics (Wang et al., 2018). By leveraging Hi-C and DNase-seq data to uncover chromatin interactions and enhancer-promoter relationships, a long-range transcriptional regulation mechanism was proposed underpinning subgenome expression coordination and partitioning.

More recently, an innovative OCEAN-C approach was applied to map genome-wide open chromatin interactions for hexaploid wheat and its tetraploid and diploid relatives (Yuan et al., 2022). By integrating OCEAN-C, ChIP- seq, ATAC-seq, and RNA-seq data, the regulatory layers of structural variations, epigenetic marks, and chromatin accessibility were jointly investigated, collectively helping to reveal the role of open chromatin interactions in shaping gene expression variation during allopolyploid evolution.

In summary, our study details changes in chromatin features genome-wide, offering insights into how allopolyploidy affects nucleosome occupancy, chromatin accessibility, and the regulatory underpinnings of expression evolution of duplicated genes. Given the broader complexity of chromatin dynamics, exploring the synergies among histone modifications, DNA methylation, enhancer-promoter interactions, and 3D chromatin organization will continue to further our understanding of the intricate web of regulatory mechanisms in shaping gene expression evolution in cotton and other allopolyploid systems.

Declarations

ACKNOWLEDGEMENTS

We thank Kenneth McCabe for Greenhouse assistance and the Research IT unit at Iowa State University (https://researchit.las.iastate.edu/) for computational support. This work was supported by grants from the National Key Research and Development Program of China (2021YFF1000100) and the National Natural Science Foundation of China (32072111). This research was funded by the National Science Foundation of United States (NSF) grants MCB-1118646 to JFW and CEG, IOS-1444532 to HWB and JZ, and IOS-1445014 to JFW.

AUTHOR CONTRIBUTIONS

JFW, HWB, and GH conceived the project and designed experiments. ERM, JPG, and CEG performed RNA-seq experiments. JAU provided ATAC-seq data. GH performed MNase-seq experiments, conducted data analyses, with inputs from P-YL, XS, DLV, SBB, JZ, and HWB for MNase-seq, from CEG for RNA-seq, from JLC for ortholog and homoeolog detection, from SO for transposable element annotation, from XX, DZ and DL for duplicated gene expression analysis. GH wrote the manuscript. CEG, HWB, and JFW revised the manuscript. All authors read and approved the final manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

References

Adams, K. L., Cronn, R., Percifield, R., & Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proceedings of the National Academy of Sciences of the United States of America, 100(8), 4649–4654.

Ahmad, K., Henikoff, S., & Ramachandran, S. (2022). Managing the steady state chromatin landscape by nucleosome dynamics. Annual Review of Biochemistry, 91, 183–195.

Andrews, A. J., & Luger, K. (2011). Nucleosome structure(s) and stability: variations on a theme. Annual Review of Biophysics, 40, 99–117.

Arrigo, N., de La Harpe, M., Litsios, G., Zozomová-Lihová, J., Španiel, S., Marhold, K., Barker, M. S., & Alvarez, N. (2016). Is hybridization driving the evolution of climatic niche in Alyssum montanum. American Journal of Botany, 103(7), 1348–1357.

Arya, G., Maitra, A., & Grigoryev, S. A. (2010). A structural perspective on the where, how, why, and what of nucleosome positioning. Journal of Biomolecular Structure & Dynamics, 27(6), 803–820.

Athey, B. D., Smith, M. F., Rankert, D. A., Williams, S. P., & Langmore, J. P. (1990). The diameters of frozen-hydrated chromatin fibers increase with DNA linker length: evidence in support of variable diameter models for chromatin. The Journal of Cell Biology, 111(3), 795–806.

Bailey, T. L., & Grant, C. E. (2021). SEA: simple enrichment analysis of motifs. In bioRxiv (p. 2021.08.23.457422). https://doi.org/10.1101/2021.08.23.457422

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. Nucleic Acids Research, 43(W1), W39–W49.

Baldi, S., Korber, P., & Becker, P. B. (2020). Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. Nature Structural & Molecular Biology, 27(2), 109–118.

Baniaga, A. E., Marx, H. E., Arrigo, N., & Barker, M. S. (2020). Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. Ecology Letters, 23(1), 68–78.

Bao, Y., Hu, G., Grover, C. E., Conover, J., Yuan, D., & Wendel, J. F. (2019). Unraveling cis and trans regulatory evolution during cotton domestication. Nature Communications, 10(1), 5399.

Barbier, J., Vaillant, C., Volff, J.-N., Brunet, F. G., & Audit, B. (2021). Coupling between sequence-mediated nucleosome organization and genome evolution. Genes, 12(6). https://doi.org/10.3390/genes12060851

Bottani, S., Zabet, N. R., Wendel, J. F., & Veitia, R. A. (2018). Gene expression dominance in allopolyploids: hypotheses and models. Trends in Plant Science. https://doi.org/10.1016/j.tplants.2018.01.002

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology, 34(5), 525–527.

Brouwer, T., Pham, C., Kaczmarczyk, A., de Voogd, W.-J., Botto, M., Vizjak, P., Mueller-Planitz, F., & van Noort, J. (2021). A critical role for linker DNA in higher-order folding of chromatin fibers. Nucleic Acids Research, 49(5), 2537–2551.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods, 10(12), 1213–1218.

Buggs, R. J., Wendel, J. F., Doyle, J. J., Soltis, D. E., Soltis, P. S., & Coate, J. E. (2014). The legacy of diploid progenitors in allopolyploid gene expression patterns. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 369(1648), 20130354.

Castro-Mondragon, J. A., Jaeger, S., Thieffry, D., Thomas-Chollier, M., & van Helden, J. (2017). RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. Nucleic Acids Research, 45(13), e119.

Chen, J., Li, E., Zhang, X., Dong, X., Lei, L., Song, W., Zhao, H., & Lai, J. (2017). Genome-wide nucleosome occupancy and organization modulates the plasticity of gene transcriptional status in maize. Molecular Plant, 10(7), 962–974.

Chen, Z., Grover, C. E., Li, P., Wang, Y., Nie, H., Zhao, Y., Wang, M., Liu, F., Zhou, Z., Wang, X., Cai, X., Wang, K., Wendel, J. F., & Hua, J. (2017). Molecular evolution of the plastid genome during diversification of the cotton genus. Molecular Phylogenetics and Evolution, 112, 268–276.

Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., Ding, M., Ye, W., Kirkbride, R. C., Jenkins, J., Plott, C., Lovell, J., Lin, Y.-M., Vaughn, R., Liu, B., Simpson, S., Scheffler, B. E., Wen, L., Saski, C. A., … Schmutz, J. (2020a). Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. Nature Genetics, 52(5), 525–533.

Chen, Z. J., Sreedasyam, A., Ando, A., Song, Q., De Santiago, L. M., Hulse-Kemp, A. M., Ding, M., Ye, W., Kirkbride, R. C., Jenkins, J., Plott, C., Lovell, J., Lin, Y.-M., Vaughn, R., Liu, B., Simpson, S., Scheffler, B. E., Wen, L., Saski, C. A., … Schmutz, J. (2020b). Genomic diversifications of five Gossypium allopolyploid species and their impact on cotton improvement. Nature Genetics, 52(5), 525–533.

Chodavarapu, R. K., Feng, S., Bernatavichute, Y. V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J. A., Kuo, F., Kim, J., Cokus, S. J., Casero, D., Bernal, M., Huijser, P., Clark, A. T., Krämer, U., Merchant, S. S., Zhang, X., Jacobsen, S. E., & Pellegrini, M. (2010). Relationship between nucleosome positioning and DNA methylation. Nature, 466(7304), 388–392.

Coate, J. E., Luciano, A. K., Seralathan, V., Minchew, K. J., Owens, T. G., & Doyle, J. J. (2012). Anatomical, biochemical, and photosynthetic responses to recent allopolyploidy in Glycine dolichocarpa (Fabaceae). American Journal of Botany, 99(1), 55–67.

Combes, M.-C., Joët, T., Stavrinides, A. K., & Lashermes, P. (2022). New cup out of old coffee: contribution of parental gene expression legacy to phenotypic novelty in coffee beans of the allopolyploid Coffea arabica L. Annals of Botany. https://doi.org/10.1093/aob/mcac041

Conover, J. L., Sharbrough, J., & Wendel, J. F. (2021). pSONIC: Ploidy-aware Syntenic Orthologous Networks Identified via Collinearity. G3 . https://doi.org/10.1093/g3journal/jkab170

Conover, J. L., & Wendel, J. F. (2022). Deleterious mutations accumulate faster in allopolyploid than diploid cotton (Gossypium) and unequally between subgenomes. Molecular Biology and Evolution, 39(2). https://doi.org/10.1093/molbev/msac024

Coughlan, J. M., Han, S., Stefanović, S., & Dickinson, T. A. (2017). Widespread generalist clones are associated with range and niche expansion in allopolyploids of Pacific Northwest Hawthorns (Crataegus L.). Molecular Ecology, 26(20), 5484–5499.

Dorrity, M. W., Alexandre, C. M., Hamm, M. O., Vigil, A.-L., Fields, S., Queitsch, C., & Cuperus, J. T. (2021). The regulatory landscape of Arabidopsis thaliana roots at single-cell resolution. Nature Communications, 12(1), 3334.

Doyle, J. J., & Coate, J. E. (2019). Polyploidy, the nucleotype, and novelty: the Impact of genome doubling on the biology of the cell. International Journal of Plant Sciences, 180(1), 1–52.

Draizen, E. J., Shaytan, A. K., Mariño-Ramírez, L., Talbert, P. B., Landsman, D., & Panchenko, A. R. (2016). HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants. Database: The Journal of Biological Databases and Curation, 2016, baw014.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792–1797.

Elliott, T. L., Muasya, M., & Bureš, P. (2022). Complex patterns of ploidy in a holocentric plant clade (Schoenus, Cyperaceae) in the Cape biodiversity hotspot. Annals of Botany. https://doi.org/10.1093/aob/mcac027

Flagel, L., Udall, J., Nettleton, D., & Wendel, J. (2008). Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution. BMC Biology, 6, 16.

Flores, O., & Orozco, M. (2011). nucleR: a package for non-parametric nucleosome positioning. In Bioinformatics (Vol. 27, Issue 15, pp. 2149–2150). https://doi.org/10.1093/bioinformatics/btr345

Fox, D. T., Soltis, D. E., Soltis, P. S., Ashman, T.-L., & Van de Peer, Y. (2020). Polyploidy: a biological force from cells to ecosystems. Trends in Cell Biology, 30(9), 688–694.

Fransz, P., & de Jong, H. (2011). From nucleosome to chromosome: a dynamic organization of genetic information. The Plant Journal: For Cell and Molecular Biology, 66(1), 4–17.

Gallagher, J. P., Grover, C. E., Hu, G., Jareczek, J. J., & Wendel, J. F. (2020). Conservation and Divergence in Duplicated Fiber Coexpression Networks Accompanying Domestication of the Polyploid L. G3 , 10(8), 2879–2892.

Gallagher, J. P., Grover, C. E., Hu, G., & Wendel, J. F. (2016). Insights into the ecology and evolution of polyploid plants through network analysis. Molecular Ecology, 25(11), 2644–2660.

Gallagher, J. P., Grover, C. E., Rex, K., Moran, M., & Wendel, J. F. (2017). A new species of cotton from Wake Atoll, Gossypium stephensii (Malvaceae). Systematic Botany, 42(1), 115–123.

Galli, M., Feng, F., & Gallavotti, A. (2020). Mapping regulatory determinants in plants. Frontiers in Genetics, 11, 591194.

Giles, K. A., & Taberlay, P. C. (2019). The role of nucleosomes in epigenetic gene regulation. In L. B. Hesson & A. L. Pritchard (Eds.), Clinical Epigenetics (pp. 87–117). Springer Singapore.

Giraud, D., Lima, O., Rousseau-Gueutin, M., Salmon, A., & Aïnouche, M. (2021). Gene and transposable element expression evolution following recent and past polyploidy events in Spartina (Poaceae). Frontiers in Genetics, 12, 589160.

Girimurugan, S. B., Liu, Y., Lung, P.-Y., Vera, D. L., Dennis, J. H., Bass, H. W., & Zhang, J. (2018). iSeg: an efficient algorithm for segmentation of genomic and epigenomic data. In BMC Bioinformatics (Vol. 19, Issue 1). https://doi.org/10.1186/s12859-018-2140-3

Gouy, M., Tannier, E., Comte, N., & Parsons, D. P. (2021). Seaview version 5: a multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation. Methods in Molecular Biology , 2231, 241–260.

Grant, C. E., & Bailey, T. L. (2021). XSTREME: Comprehensive motif analysis of biological sequence datasets. bioRxiv, 2021.09.02.458722.

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics , 27(7), 1017–1018.

Grossman, S. R., Engreitz, J., Ray, J. P., Nguyen, T. H., Hacohen, N., & Lander, E. S. (2018). Positional specificity of different transcription factor classes within enhancers. Proceedings of the National Academy of Sciences of the United States of America, 115(30), E7222–E7230.

Grover, C. E., Gallagher, J. P., Szadkowski, E. P., Yoo, M. J., Flagel, L. E., & Wendel, J. F. (2012). Homoeolog expression bias and expression level dominance in allopolyploids. The New Phytologist, 196(4), 966–971.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology, 59(3), 307–321.

Han, J., Lopez-Arredondo, D., Yu, G., Wang, Y., Wang, B., Wall, S. B., Zhang, X., Fang, H., Barragán-Rosillo, A. C., Pan, X., Jiang, Y., Chen, J., Zhang, H., Zhou, B.-L., Herrera-Estrella, L., Zhang, B., & Wang, K. (2022). Genome-wide chromatin accessibility analysis unveils open chromatin convergent evolution during polyploidization in cotton. Proceedings of the National Academy of Sciences, 119(44), e2209743119.

Han, J., Wang, P., Wang, Q., Lin, Q., Chen, Z., Yu, G., Miao, C., Dao, Y., Wu, R., Schnable, J. C., Tang, H., & Wang, K. (2020). Genome-wide characterization of DNase I-hypersensitive sites and cold response regulatory landscapes in grasses. The Plant Cell, 32(8), 2457–2473.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., & Glass, C. K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. In Molecular Cell (Vol. 38, Issue 4, pp. 576–589). https://doi.org/10.1016/j.molcel.2010.05.004

Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M., & Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proceedings of the National Academy of Sciences of the United States of America, 108(45), 18318–18323.

Heslop-Harrison, J. S. P., Schwarzacher, T., & Liu, Q. (2022). Polyploidy: its consequences and enabling role in plant diversification and evolution. Annals of Botany. https://doi.org/10.1093/aob/mcac132

Hofmeister, B. T., Lee, K., Rohr, N. A., Hall, D. W., & Schmitz, R. J. (2017). Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation. Genome Biology, 18(1), 155.

Huang, G., Huang, J.-Q., Chen, X.-Y., & Zhu, Y.-X. (2021). Recent advances and future perspectives in cotton research. Annual Review of Plant Biology, 72, 437–462.

Huang, G., Wu, Z., Percy, R. G., Bai, M., Li, Y., Frelichowski, J. E., Hu, J., Wang, K., Yu, J. Z., & Zhu, Y. (2020). Genome sequence of Gossypium herbaceum and genome updates of Gossypium arboreum and Gossypium hirsutum provide insights into cotton A-genome evolution. In Nature Genetics (Vol. 52, Issue 5, pp. 516–524). https://doi.org/10.1038/s41588-020-0607-4

Hu, G., Grover, C. E., Yuan, D., Dong, Y., Miller, E., Conover, J. L., & Wendel, J. F. (2021). Evolution and diversity of the cotton genome. In M. U. Rahman, Y. Zafar, & T. Zhang (Eds.), Cotton Precision Breeding. Springer International Publishing.

Hu, G., Hovav, R., Grover, C. E., Faigenboim-Doron, A., Kadmon, N., Page, J. T., Udall, J. A., & Wendel, J. F. (2016). Evolutionary conservation and divergence of gene coexpression networks in Gossypium (cotton) seeds. Genome Biology and Evolution, 8(12), 3765–3783.

Hu, G., Koh, J., Yoo, M.-J., Chen, S., & Wendel, J. F. (2015). Gene-expression novelty in allopolyploid cotton: a proteomic perspective. Genetics, 200(1), 91–104.

Hu, G., Koh, J., Yoo, M.-J., Grupp, K., Chen, S., & Wendel, J. F. (2013). Proteomic profiling of developing cotton fibers from wild and domesticated Gossypium barbadense. The New Phytologist, 200(2), 570–582.

Hu, G., Koh, J., Yoo, M.-J., Pathak, D., Chen, S., & Wendel, J. F. (2014). Proteomics profiling of fiber development and domestication in upland cotton (Gossypium hirsutum L.). Planta, 240(6), 1237–1251.

Hu, G., & Wendel, J. F. (2019). Cis-trans controls and regulatory novelty accompanying allopolyploidization. The New Phytologist, 221(4), 1691–1700.

Initiative, O. T. P. T., & One Thousand Plant Transcriptomes Initiative. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. In Nature (Vol. 574, Issue 7780, pp. 679–685). https://doi.org/10.1038/s41586-019-1693-2

Jackson, S. A. (2017). Epigenomics: dissecting hybridization and polyploidization. Genome Biology, 18(1), 117.

Jiang, J. (2015). The “dark matter”in the plant genomes: non-coding and unannotated DNA sequences associated with open chromatin. Current Opinion in Plant Biology, 24, 17–23.

Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., Tomsho, L. P., Hu, Y., Liang, H., Soltis, P. S., Soltis, D. E., Clifton, S. W., Schlarbaum, S. E., Schuster, S. C., Ma, H., Leebens-Mack, J., & dePamphilis, C. W. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature, 473(7345), 97–100.

Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., & Gao, G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Research, 45(D1), D1040–D1045.

Jordan, K. W., He, F., de Soto, M. F., Akhunova, A., & Akhunov, E. (2020). Differential chromatin accessibility landscape reveals structural and functional features of the allopolyploid wheat chromosomes. In Genome Biology (Vol. 21, Issue 1). https://doi.org/10.1186/s13059-020-02093-1

Jordan, W. T., & Schmitz, R. J. (2016). The shocking consequences of hybrid epigenomes [Review of The shocking consequences of hybrid epigenomes]. Genome Biology, 17, 85.

Kaplan, N., Moore, I. K., Fondufe-Mittendorf, Y., Gossett, A. J., Tillo, D., Field, Y., LeProust, E. M., Hughes, T. R., Lieb, J. D., Widom, J., & Segal, E. (2009). The DNA-encoded nucleosome organization of a eukaryotic genome. Nature, 458(7236), 362–366.

Kawakatsu, T., Huang, S.-S. C., Jupe, F., Sasaki, E., Schmitz, R. J., Urich, M. A., Castanon, R., Nery, J. R., Barragan, C., He, Y., Chen, H., Dubin, M., Lee, C.-R., Wang, C., Bemm, F., Becker, C., O’Neil, R., O’Malley, R. C., Quarless, D. X., … Ecker. (2016). Epigenomic diversity in a global collection of Arabidopsis thaliana accessions. Cell, 166(2), 492–505.

Klein, D. C., & Hainer, S. J. (2020). Genomic methods in profiling DNA accessibility and factor localization. Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology, 28(1), 69–85.

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. Nature Reviews. Genetics, 20(4), 207–220.

Knight, C. A., & Beaulieu, J. M. (2008). Genome size scaling through phenotype space. Annals of Botany, 101(6), 759–766.

Kolde, R. (2017). Pheatmap: pretty heatmaps. 2015. R Package Version, 1(10).

Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science, 184(4139), 868–871.

Krueger, F. (2012). Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. URL Http://www. Bioinformatics. Babraham. Ac. Uk/projects/trim_galore/. (Date of Access: 28/04/2016).

Krueger, F. (2015). Trim galore. A Wrapper Tool around Cutadapt and FastQC to Consistently Apply Quality and Adapter Trimming to FastQ Files.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., … Snyder, M. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Research, 22(9), 1813–1831.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods, 9(4), 357–359.

Lantermann, A. B., Straub, T., Strålfors, A., Yuan, G.-C., Ekwall, K., & Korber, P. (2010). Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. Nature Structural & Molecular Biology, 17(2), 251–257.

Leitch, A. R., & Leitch, I. J. (2008). Genomic plasticity and the diversity of polyploid Ppants. Science, 320(5875), 481–483.

Levin, D. A. (1983). Polyploidy and novelty in flowering plants. The American Naturalist, 122(1), 1–25.

Li, G., Liu, S., Wang, J., He, J., Huang, H., Zhang, Y., & Xu, L. (2014). ISWI proteins participate in the genome-wide nucleosome distribution in Arabidopsis. The Plant Journal: For Cell and Molecular Biology, 78(4), 706–714.

Liu, M.-J., Seddon, A. E., Tsai, Z. T.-Y., Major, I. T., Floer, M., Howe, G. A., & Shiu, S.-H. (2015). Determinants of nucleosome positioning and their influence on plant gene expression. Genome Research, 25(8), 1182–1195.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15(12), 550.

Luger, K., Rechsteiner, T. J., Flaus, A. J., Waye, M. M., & Richmond, T. J. (1997). Characterization of nucleosome core particles containing histone proteins made in bacteria. Journal of Molecular Biology, 272(3), 301–311.

Lun, A. T. L., & Smyth, G. K. (2016). csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. Nucleic Acids Research, 44(5), e45.

Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M., & Schmitz, R. J. (2017a). Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Research, 45(6), e41–e41.

Lu, Z., Hofmeister, B. T., Vollmers, C., DuBois, R. M., & Schmitz, R. J. (2017b). Combining ATAC-seq with nuclei sorting for discovery of cis-regulatory regions in plant genomes. Nucleic Acids Research, 45(6), e41.

Lu, Z., Marand, A. P., Ricci, W. A., Ethridge, C. L., Zhang, X., & Schmitz, R. J. (2019). The prevalence, evolution and chromatin signatures of plant regulatory elements. Nature Plants. https://doi.org/10.1038/s41477-019-0548-z

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. In EMBnet.journal (Vol. 17, Issue 1, p. 10). https://doi.org/10.14806/ej.17.1.200

Mata, J. K., Martin, S. L., & Smith, T. W. (2023). Global biodiversity data suggest allopolyploid plants do not occupy larger ranges or harsher conditions compared with their progenitors. Ecology and Evolution, 13(8), e10231.

McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. BMC Bioinformatics, 11, 165.

Mieczkowski, J., Cook, A., Bowman, S. K., Mueller, B., Alver, B. H., Kundu, S., Deaton, A. M., Urban, J. A., Larschan, E., Park, P. J., Kingston, R. E., & Tolstorukov, M. Y. (2016). MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. Nature Communications, 7, 11485.

Mishra, M. K. (1997). Stomatal characteristics at different ploidy levels in Coffea L. Annals of Botany, 80(5), 689–692.

Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. D., Li, Q., Rohr, N. A., Rambani, A., Burke, J. M., Udall, J. A., Egesi, C., Schmutz, J., Grimwood, J., Jackson, S. A., Springer, N. M., & Schmitz, R. J. (2016). Widespread natural variation of DNA methylation within angiosperms. Genome Biology, 17(1), 194.

Olsen, K. M., & Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. Annual Review of Plant Biology, 64, 47–70.

Orr-Weaver, T. L. (2015). When bigger is better: the role of polyploidy in organogenesis. Trends in Genetics: TIG, 31(6), 307–315.

Otto, S. P. (2007). The evolutionary consequences of polyploidy. Cell, 131(3), 452–462.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biology, 20(1), 275.

Page, J. T., Gingle, A. R., & Udall, J. A. (2013). PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. G3 , 3(3), 517–525.

Parshuram, Z. A., Harrison, T. L., Simonsen, A. K., Stinchcombe, J. R., & Frederickson, M. E. (2022). Nonsymbiotic legumes are more invasive, but only if polyploid. The New Phytologist. https://doi.org/10.1111/nph.18579

Parvathaneni, R. K., Bertolini, E., Shamimuzzaman, M., Vera, D. L., Lung, P.-Y., Rice, B. R., Zhang, J., Brown, P. J., Lipka, A. E., Bass, H. W., & Eveland, A. L. (2020). The regulatory landscape of early maize inflorescence development. Genome Biology, 21(1), 165.

Pass, D. A., Sornay, E., Marchbank, A., Crawford, M. R., Paszkiewicz, K., Kent, N. A., & Murray, J. A. H. (2017). Genome-wide chromatin mapping with size resolution reveals a dynamic sub-nucleosomal landscape in Arabidopsis. PLoS Genetics, 13(9), e1006988.

Paterson, A. H., Wendel, J. F., Gundlach, H., Guo, H., Jenkins, J., Jin, D., Llewellyn, D., Showmaker, K. C., Shu, S., Udall, J., Yoo, M.-J., Byers, R., Chen, W., Doron-Faigenboim, A., Duke, M. V., Gong, L., Grimwood, J., Grover, C., Grupp, K., … Schmutz, J. (2012). Repeated polyploidization of Gossypium genomes and the evolution of spinnable cotton fibres. Nature, 492(7429), 423–427.

Pei, L., Li, G., Lindsey, K., Zhang, X., & Wang, M. (2021). Plant 3D genomics: the exploration and application of chromatin organization. The New Phytologist, 230(5), 1772–1786.

Probst, A. V., Desvoyes, B., & Gutierrez, C. (2020). Similar yet critically different: the distribution, dynamics and function of histone variants. Journal of Experimental Botany, 71(17), 5191–5204.

Qiu, Z., Li, R., Zhang, S., Wang, K., Xu, M., Li, J., Du, Y., Yu, H., & Cui, X. (2016). Identification of regulatory DNA elements using genome-wide mapping of DNase I hypersensitive sites during tomato fruit development. Molecular Plant, 9(8), 1168–1182.

Quinlan, A. R. (2014). BEDTools: The swiss-army tool for genome feature analysis. Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.], 47, 11.12.1–34.

Radman-Livaja, M., & Rando, O. J. (2010). Nucleosome positioning: How is it established, and why does it matter? In Developmental Biology (Vol. 339, Issue 2, pp. 258–266). https://doi.org/10.1016/j.ydbio.2009.06.012

Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A., & Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Research, 42(Web Server issue), W187–W191.

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research, 44(W1), W160–W165.

Ramsey, J., & Schemske, D. W. (2002). Neopolyploidy in flowering plants. Annual Review of Ecology and Systematics, 33(1), 589–639.

Rapp, R. A., Udall, J. A., & Wendel, J. F. (2009). Genomic expression dominance in allopolyploids. BMC Biology, 7, 18.

Renny-Byfield, S., & Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. American Journal of Botany, 101(10), 1711–1725.

Reske, J. J., Wilson, M. R., & Chandler, R. L. (2020). ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. Epigenetics & Chromatin, 13(1), 22.

Reynoso, M. A., Borowsky, A. T., Pauluzzi, G. C., Yeung, E., Zhang, J., Formentin, E., Velasco, J., Cabanlit, S., Duvenjian, C., Prior, M. J., Akmakjian, G. Z., Deal, R. B., Sinha, N. R., Brady, S. M., Girke, T., & Bailey-Serres, J. (2022). Gene regulatory networks shape developmental plasticity of root cell types under water extremes in rice. Developmental Cell, 57(9), 1177–1192.e6.

Ricci, W. A., Lu, Z., Ji, L., Marand, A. P., Ethridge, C. L., Murphy, N. G., Noshay, J. M., Galli, M., Mejía-Guerra, M. K., Colomé-Tatché, M., Johannes, F., Rowley, M. J., Corces, V. G., Zhai, J., Scanlon, M. J., Buckler, E. S., Gallavotti, A., Springer, N. M., Schmitz, R. J., & Zhang, X. (2019). Widespread long-range cis-regulatory elements in the maize genome. Nature Plants, 5(12), 1237–1249.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. Nature Biotechnology, 29(1), 24–26.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics , 26(1), 139–140.

Rodgers-Melnick, E., Vera, D. L., Bass, H. W., & Buckler, E. S. (2016). Open chromatin reveals the functional maize genome. Proceedings of the National Academy of Sciences of the United States of America, 113(22), E3177–E3184.

Ruprecht, C., Lohaus, R., Vanneste, K., Mutwil, M., Nikoloski, Z., Van de Peer, Y., & Persson, S. (2017). Revisiting ancestral polyploidy in plants. Science Advances, 3(7), e1603195.

Savadel, S. D., Hartwig, T., Turpin, Z. M., Vera, D. L., Lung, P.-Y., Sui, X., Blank, M., Frommer, W. B., Dennis, J. H., Zhang, J., & Bass, H. W. (2021). The native cistrome and sequence motif families of the maize ear. PLoS Genetics, 17(8), e1009689.

Schmitz, R. J., Schultz, M. D., Lewsey, M. G., O'Malley, R. C., Urich, M. A., Libiger, O., Schork, N. J., & Ecker, J. R. (2011). Transgenerational epigenetic instability is a source of novel methylation variants. Science, 334(6054), 369–373.

Schnable, J. C., Springer, N. M., & Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences of the United States of America, 108(10), 4069–4074.

Segal, E., & Widom, J. (2009). What controls nucleosome positions? Trends in Genetics: TIG, 25(8), 335–343.

Shan, S., Boatwright, J. L., Liu, X., Chanderbali, A. S., Fu, C., Soltis, P. S., & Soltis, D. E. (2020). Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid Tragopogon miscellus (Asteraceae). Frontiers in Genetics, 11, 888.

Soltis, P. S., & Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. Current Opinion in Plant Biology, 30, 159–165.

Soman, A., Wong, S. Y., Korolev, N., Surya, W., Lattmann, S., Vogirala, V. K., Chen, Q., Berezhnoy, N. V., van Noort, J., Rhodes, D., & Nordenskiöld, L. (2022). Columnar structure of human telomeric chromatin. Nature, 609(7929), 1048–1055.

Song, Q., Zhang, T., Stelly, D. M., & Chen, Z. J. (2017). Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biology, 18(1), 99.

Springer, N. M., & Schmitz, R. J. (2017). Exploiting induced and natural epigenetic variation for crop improvement. Nature Reviews. Genetics, 18(9), 563–575.

Stebbins, G. L. (1940). The significance of polyploidy in plant evolution. The American Naturalist, 74(750), 54–66.

Stewart, J. M., Craven, L. A., Brubaker, C., & Wendel, J. F. (2015). Gossypium anapoides (Malvaceae), a new species from western Australia. In Novon: A Journal for Botanical Nomenclature (Vol. 23, Issue 4, pp. 447–451). https://doi.org/10.3417/2007140

Sugiyama, S.-I. (2005). Polyploidy and cellular mechanisms changing leaf size: comparison of diploid and autotetraploid populations in two species of Lolium. Annals of Botany, 96(5), 931–938.

Sullivan, A. M., Bubb, K. L., Sandstrom, R., Stamatoyannopoulos, J. A., & Queitsch, C. (2015). DNase I hypersensitivity mapping, genomic footprinting, and transcription factor networks in plants. Current Plant Biology, 3-4, 40–47.

Talbert, P. B., Meers, M. P., & Henikoff, S. (2019). Old cogs, new tricks: the evolution of gene expression in a chromatin context. Nature Reviews. Genetics, 20(5), 283–297.

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. Molecular Biology and Evolution, 38(7), 3022–3027.

Teves, S. S., & Henikoff, S. (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes & Development, 25(22), 2387–2397.

Tommerup, H., Dousmanis, A., & de Lange, T. (1994). Unusual chromatin in human telomeres. Molecular and Cellular Biology, 14(9), 5777–5785.

Tsompana, M., & Buck, M. J. (2014). Chromatin accessibility: a window into the genome. Epigenetics & Chromatin, 7(1), 33.

Turpin, Z. M., Vera, D. L., Savadel, S. D., Lung, P.-Y., Wear, E. E., Mickelson-Young, L., Thompson, W. F., Hanley-Bowdoin, L., Dennis, J. H., Zhang, J., & Bass, H. W. (2018). Chromatin structure profile data from DNS-seq: differential nuclease sensitivity mapping of four reference tissues of B73 maize (Zea mays L). In Data in Brief (Vol. 20, pp. 358–363). https://doi.org/10.1016/j.dib.2018.08.015

Vainshtein, Y., Rippe, K., & Teif, V. B. (2017). NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data. BMC Genomics, 18(1), 158.

Valouev, A., Johnson, S. M., Boyd, S. D., Smith, C. L., Fire, A. Z., & Sidow, A. (2011). Determinants of nucleosome organization in primary human cells. Nature, 474(7352), 516–520.

Van de Peer, Y., Ashman, T.-L., Soltis, P. S., & Soltis, D. E. (2021). Polyploidy: an evolutionary and ecological force in stressful times. The Plant Cell, 33(1), 11–26.

Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. Nature Reviews. Genetics, 18(7), 411–424.

Vera, D. L., Madzima, T. F., Labonne, J. D., Alam, M. P., Hoffman, G. G., Girimurugan, S. B., Zhang, J., McGinnis, K. M., Dennis, J. H., & Bass, H. W. (2014). Differential nuclease sensitivity profiling of chromatin reveals biochemical footprints coupled to gene expression and functional DNA elements in maize. The Plant Cell, 26(10), 3883–3893.

Vershinin, A. V., & Heslop-Harrison, J. S. (1998). Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. Plant Molecular Biology, 36(1), 149–161.

Viot, C. R., & Wendel, J. F. (2023). Evolution of the cotton genus, Gossypium, and its domestication in the Americas. Critical Reviews in Plant Sciences, 42(1), 1–33.

Visger, C. J., Wong, G. K.-S., Zhang, Y., Soltis, P. S., & Soltis, D. E. (2019). Divergent gene expression levels between diploid and autotetraploid Tolmiea relative to the total transcriptome, the cell, and biomass. American Journal of Botany, 106(2), 280–291.

Voong, L. N., Xi, L., Wang, J.-P., & Wang, X. (2017). Genome-wide mapping of the nucleosome landscape by micrococcal nuclease and chemical mapping. Trends in Genetics: TIG, 33(8), 495–507.

Wang, G., Zhou, N., Chen, Q., Yang, Y., Yang, Y., & Duan, Y. (2021). Gradual genome size evolution and polyploidy in Allium from the Qinghai-Tibetan Plateau. Annals of Botany. https://doi.org/10.1093/aob/mcab155

Wang, M., Tu, L., Lin, M., Lin, Z., Wang, P., Yang, Q., Ye, Z., Shen, C., Li, J., Zhang, L., Zhou, X., Nie, X., Li, Z., Guo, K., Ma, Y., Huang, C., Jin, S., Zhu, L., Yang, X., … Zhang, X. (2017). Asymmetric subgenome selection and cis-regulatory divergence during cotton domestication. Nature Genetics, 49(4), 579–587.

Wang, M., Wang, P., Lin, M., Ye, Z., Li, G., Tu, L., Shen, C., Li, J., Yang, Q., & Zhang, X. (2018). Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. Nature Plants, 4(2), 90–97.

Weintraub, H., & Groudine, M. (1976). Chromosomal subunits in active genes have an altered conformation. Science, 193(4256), 848–856.

Wendel, J. F. (2015). The wondrous cycles of polyploidy in plants. American Journal of Botany, 102(11), 1753–1756.

Wendel, J. F., & Albert, V. A. (1992). Phylogenetics of the cotton genus (Gossypium): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. Systematic Botany, 17(1), 115.

Wendel, J. F., Brubaker, C. L., & Seelanan, T. (2010). The origin and evolution of Gossypium. In J. M. Stewart, D. M. Oosterhuis, J. J. Heitholt, & J. R. Mauney (Eds.), Physiology of Cotton (pp. 1–18). Springer Netherlands.

Wendel, J. F., & Cronn, R. C. (2003). Polyploidy and the evolutionary history of cotton. Advances in Agronomy, 78, 139–186.

Wendel, J. F., & Grover, C. E. (2015). Taxonomy and evolution of the cotton genus, Gossypium. In D. D. Fang & R. G. Percy (Eds.), Cotton (2nd ed., pp. 25–42). American Society of Agronomy.

Wendel, J. F., Lisch, D., Hu, G., & Mason, A. S. (2018). The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Current Opinion in Genetics & Development, 49, 1–7.

Wendel, J. F., Olson, P. D., & Stewart, J. M. (1989). Genetic diversity, introgression, and independent domestication of Old World cultivated cottons. American Journal of Botany, 76(12), 1795–1806.

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. Nature, 430(6995), 85–88.

Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R., & Elgin, S. C. (1979). The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. Cell, 16(4), 797–806.

Wu, C., Wong, Y. C., & Elgin, S. C. (1979). The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. Cell, 16(4), 807–814.

Wu, Y., Zhang, W., & Jiang, J. (2014). Genome-wide nucleosome positioning is orchestrated by genomic regions associated with DNase I hypersensitivity in rice. PLoS Genetics, 10(5), e1004378.

Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J., & Wang, J.-P. (2010). Predicting nucleosome positioning using a duration Hidden Markov Model. BMC Bioinformatics, 11, 346.

Yoo, M. J., Liu, X., Pires, J. C., Soltis, P. S., & Soltis, D. E. (2014). Nonadditive gene expression in polyploids. Annual Review of Genetics, 48, 485–517.

Yoo, M. J., Szadkowski, E., & Wendel, J. F. (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity, 110(2), 171–180.

Yoo, M. J., & Wendel, J. F. (2014). Comparative evolutionary and developmental dynamics of the cotton (Gossypium hirsutum) fiber transcriptome. PLoS Genetics, 10(1), e1004073.

Yuan, J., Sun, H., Wang, Y., Li, L., Chen, S., Jiao, W., Jia, G., Wang, L., Mao, J., Ni, Z., Wang, X., & Song, Q. (2022). Open chromatin interaction maps reveal functional regulatory elements and chromatin architecture variations during wheat evolution. Genome Biology, 23(1), 34.

Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics , 31(14), 2382–2383.

Yu, J., Jung, S., Cheng, C.-H., Ficklin, S. P., Lee, T., Zheng, P., Jones, D., Percy, R. G., & Main, D. (2014). CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Research, 42(Database issue), D1229–D1236.

Zentner, G. E., & Henikoff, S. (2012). Surveying the epigenomic landscape, one base at a time. Genome Biology, 13(10), 250–250.

Zhang, K., Wang, X., & Cheng, F. (2019). Plant polyploidy: origin, evolution, and its influence on crop domestication. Horticultural Plant Journal, 5(6), 231–239.

Zhang, K., Zhao, X., Zhao, Y., Zhang, Z., Liu, Z., Liu, Z., Yu, Y., Li, J., Ma, Y., Dong, Y., Pang, X., Jin, X., Li, N., Liu, B., Wendel, J. F., Zhai, J., Long, Y., Wang, T., & Gong, L. (2023). Cell type-specific cytonuclear co-evolution in three allopolyploid plant species. Proceedings of the National Academy of Sciences of the United States of America, in press.

Zhang, T., Zhang, W., & Jiang, J. (2015). Genome-wide nucleosome occupancy and positioning and their impact on gene expression and evolution in plants. In Plant Physiology (Vol. 168, Issue 4, pp. 1406–1416). https://doi.org/10.1104/pp.15.00125

Zhang, W., & Jiang, J. (2018). Application of MNase-seq in the global mapping of nucleosome positioning in plants. Methods in Molecular Biology , 1830, 353–366.

Zhang, W., Wu, Y., Schnable, J. C., Zeng, Z., Freeling, M., Crawford, G. E., & Jiang, J. (2012). High-resolution mapping of open chromatin in the rice genome. Genome Research, 22(1), 151–162.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biology, 9(9), R137.

Zhao, H., Zhang, W., Chen, L., Wang, L., Marand, A. P., Wu, Y., & Jiang, J. (2018). Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. Plant Physiology, 176(4), 2789–2803.

Zhao, H., Zhang, W., Zhang, T., Lin, Y., Hu, Y., Fang, C., & Jiang, J. (2020). Genome-wide MNase hypersensitivity assay unveils distinct classes of open chromatin associated with H3K27me3 and DNA methylation in Arabidopsis thaliana. In Genome Biology (Vol. 21, Issue 1). https://doi.org/10.1186/s13059-020-1927-5

Zhao, L., Yang, Y.-Y., Qu, X.-J., Ma, H., Hu, Y., Li, H.-T., Yi, T.-S., & Li, D.-Z. (2022). Phylotranscriptomic analyses reveal multiple whole-genome duplication events, the history of diversification and adaptations in the Araceae. Annals of Botany. https://doi.org/10.1093/aob/mcac062

Zhao, X. P., Si, Y., Hanson, R. E., Crane, C. F., Price, H. J., Stelly, D. M., Wendel, J. F., & Paterson, A. H. (1998). Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. Genome Research, 8(5), 479–492.

Zheng, D., Ye, W., Song, Q., Han, F., Zhang, T., & Chen, Z. J. (2016). Histone modifications define expression bias of homoeologous genomes in allotetraploid cotton. Plant Physiology, 172(3), 1760–1771.

Zlatanova, J., Seebart, C., & Tomschik, M. (2008). The linker-protein network: control of nucleosomal DNA accessibility. Trends in Biochemical Sciences, 33(6), 247–253.

Tables

Table 1 is available in the supplementary files section.

Figures

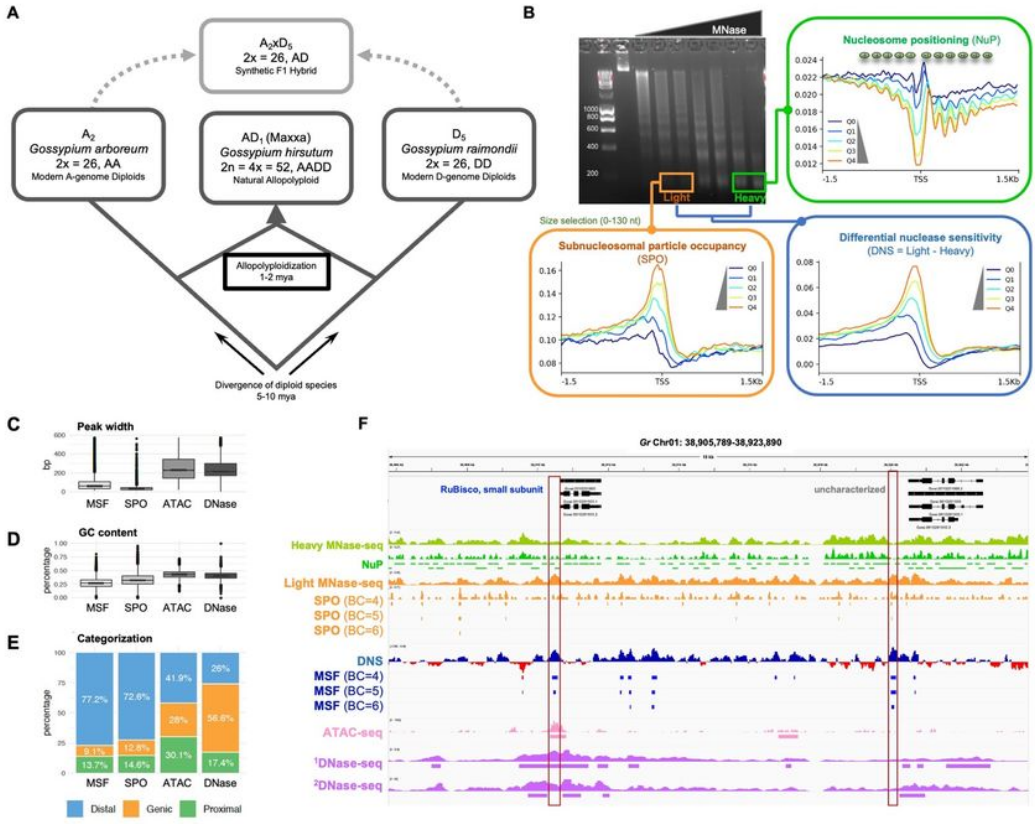


Figure 1

Studying chromatin structure evolution in diploid and allopolyploid cotton.

A. Four *Gossypium* genotypes were used in this study: a natural allopolyploid, *G. hirsutum* cultivar Acala Maxxa (AD_1); the model A- and D-genome diploid progenitors - *G. arboreum* accession A_2-101 (A_2) cultivar and *G. raimondii* (D_5); and their corresponding interspecific diploid F_1 hybrid (A_2×D_5).

B. The technique of DNS-seq was used to profile various chromatin features, including nucleosome positioning (NuP), subnucleosomal particle occupancy (SPO), and differential nuclease sensitivity (DNS). The agarose gel image shows nucleosomal DNA laddering from MNase digestions, where 5.6 U/mL and 0.4 U/mL were selected for heavy and light digestion, respectively. For each chromatin feature, aggregate plots are shown spanning ± 1.5 kb around the transcription start site (TSS) and binned by five gene expression level groups, where Q1 to Q4 represent increasing expression quantiles, and Q0 represents the group of non-expressed genes.

C-E. Accessible chromatin regions (ACRs) were compared between the analyses of MSF, SPO, ATAC-seq, and DNase-seq (see Table S6), in terms of peak width (**C**), GC content (**D**), and categorization relative to nearest genes (**E**). Genic - ACRs are located within, or overlapped with, gene regions; Proximal - within 2 kb regions flanking genes; Distal - outside 2 kb regions flanking genes.

F. A representative 18 kb region from D_5 chromosome 1 showing a comparison of chromatin profiles by DNS-seq, ATAC-seq, and DNase-seq. Two leaf DNase-seq datasets were included: [1](Han et al., 2022) and [2](Wang et al., 2018). The gene *Gorai.001G201800*, encoding the small subunit of the chloroplast photosynthetic enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), was the most expressed gene in D_5. Identified promoter ACRs are marked by red boxes.

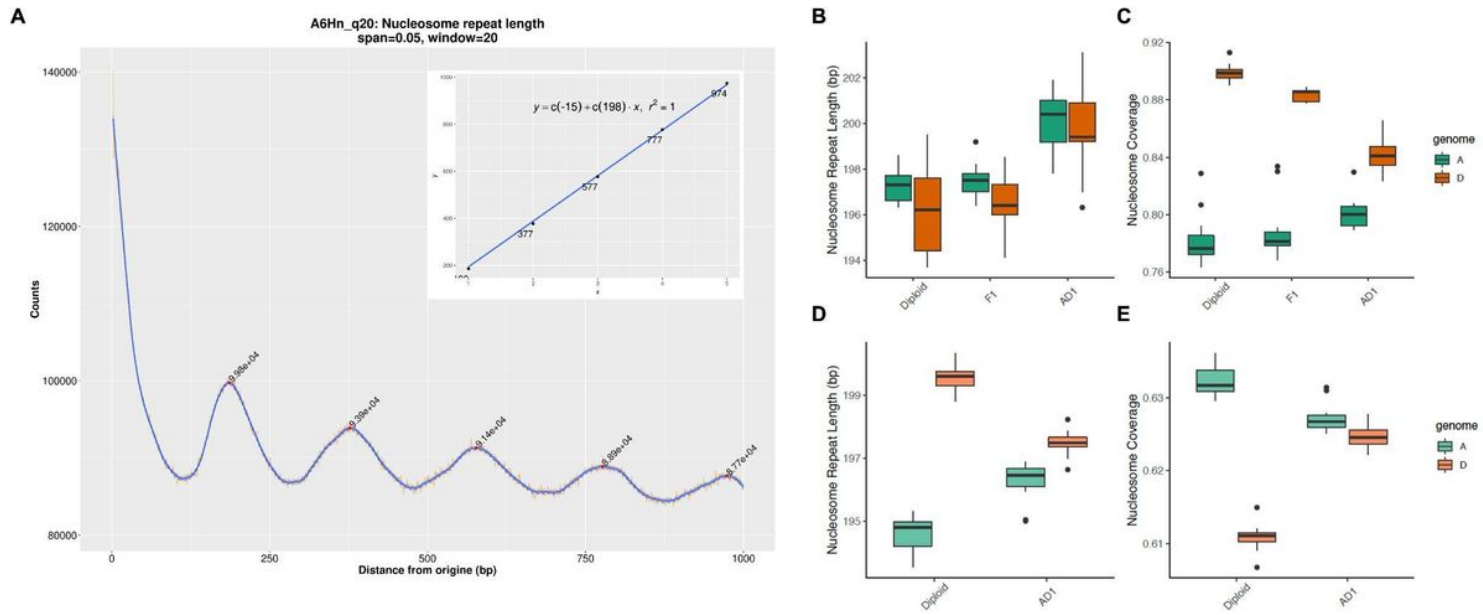


Figure 2

Comparing nucleosome organization in diploid, hybrid, and allopolyploid cottons. **A**. Nucleosome phasogram exhibits a wave-like pattern of distances between neighboring nucleosome centers. Inset presents a linear fit to the positions of the phase peaks, where the slope represents the estimated nucleosome repeat length (NRL) of 198 bp in the exemplar, A_2. **B**. Estimated NRL by phasogram across diploid and polyploid cotton genomes. **C**. Estimated nucleosome coverage (NC) based on the nucleosome positioning profiled by MNase-seq under heavy digestive conditions. **D-E**. Predicted NRL and NC based on reference genome sequence, respectively.

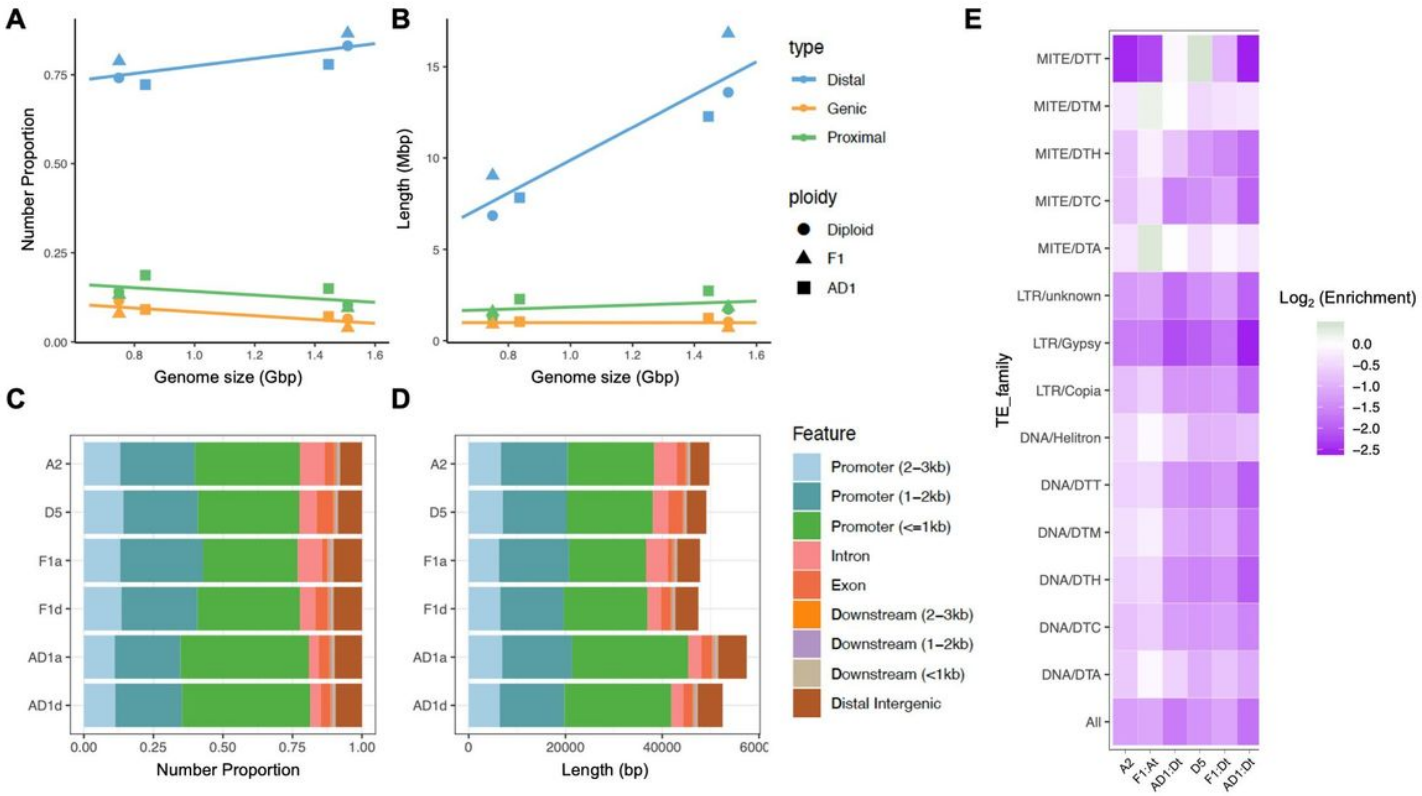


Figure 3

Comparing ACRs in diploid, hybrid, and allopolyploid cotton.

A-**B**. ACRs were categorized as genic (gACRs), proximal (pACRs), or distal ACRs (dACRs). Their relative proportions (**A**) and total lengths (**B**) are presented (Y-axis) against corresponding genome sizes (X-axis), with a linear regression trendline plotted per category. The reference genome sizes used are: A_2 = F_1:At = 1.51 Gb; D_5 = F_1:Dt = 0.75 Gb; AD_1:At = 1.45 Gb; AD_1:Dt = 0.84 Gb.

C-**D**. Parsed categorization of gACRs and pACRs using detailed genomic annotations from ChIPseeker, displayed as peak proportions (**C**) and total lengths (**D**). ANOVA followed by Tukey's post hoc test found significant increases in both proportion and total length of pACRs within the AD_1 1 kb promoter regions (P < 0.05).

E. Heatmap of ACR presence in TEs. Enrichment scores were calculated as the $\log_2$-transformed fold changes of observed versus expected (estimated from 1000 permutations) mean ACR proportions within TE superfamilies.

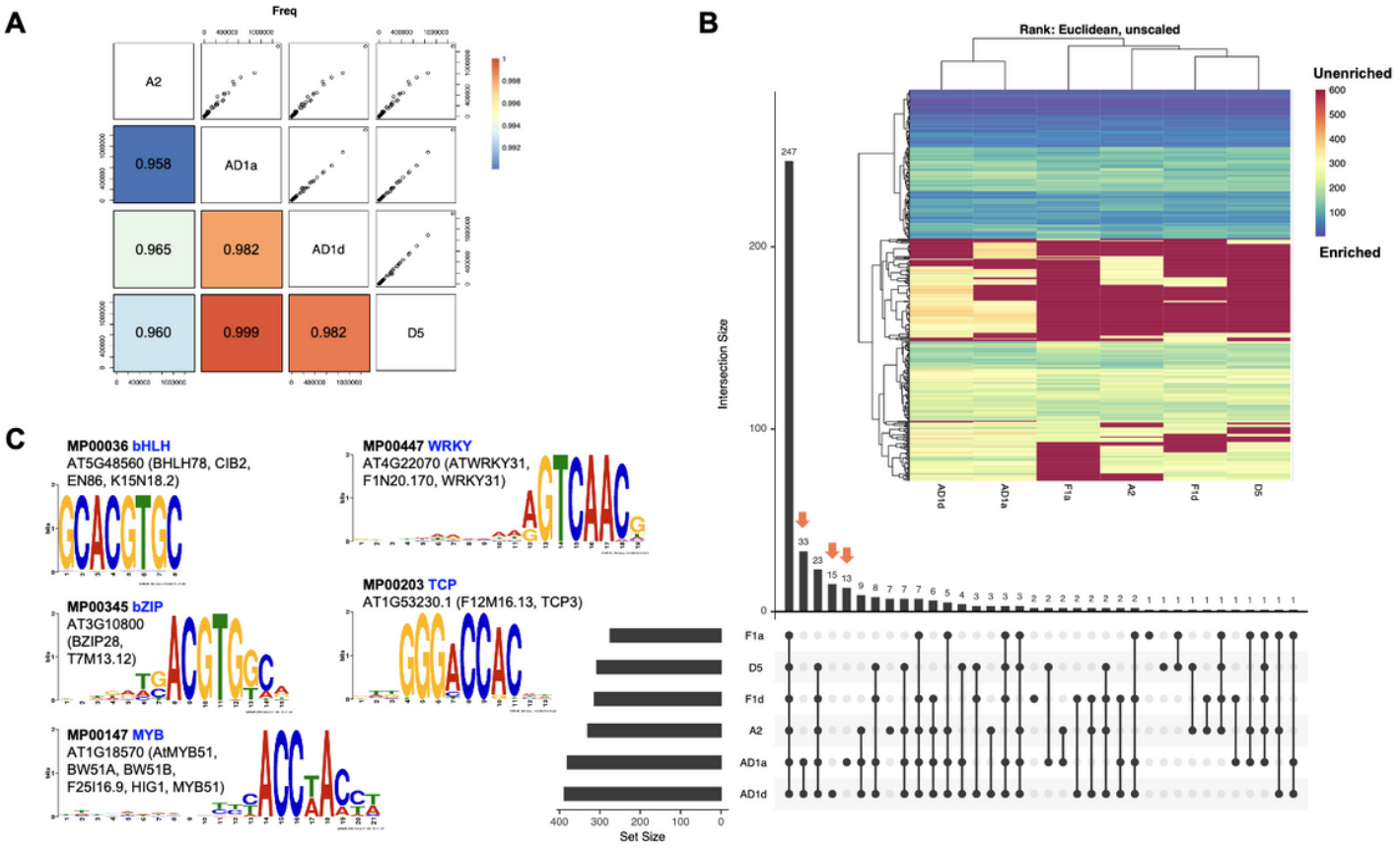


Figure 4

Motif analysis of pACRs in 1 kb promoters. **A**. Comparing background promoter sequences based on FIMO scanning for known motifs from plantTFDB v5.0 (Jin et al., 2017). Upper triangle of the matrix shows the scatter plots of motif frequency, and the lower triangle presents the pairwise Pearson correlation coefficients of motif frequency. **B**. A union of 423 enriched motifs within the1 kb promoter pACRs by AME. The UpSetplot presents the intersection of motif sets, with AD_1-specific sets marked by orange arrows. For each (sub)genome, the enrichment ranking of motifs was used for clustering and heatmap visualization, i.e. the lower ranking indicates more enrichments. A ranking score of 600 (dark red) was assigned to unenriched motifs in the corresponding genome. **C**. Top 5 most enriched AD_1-specific motifs.

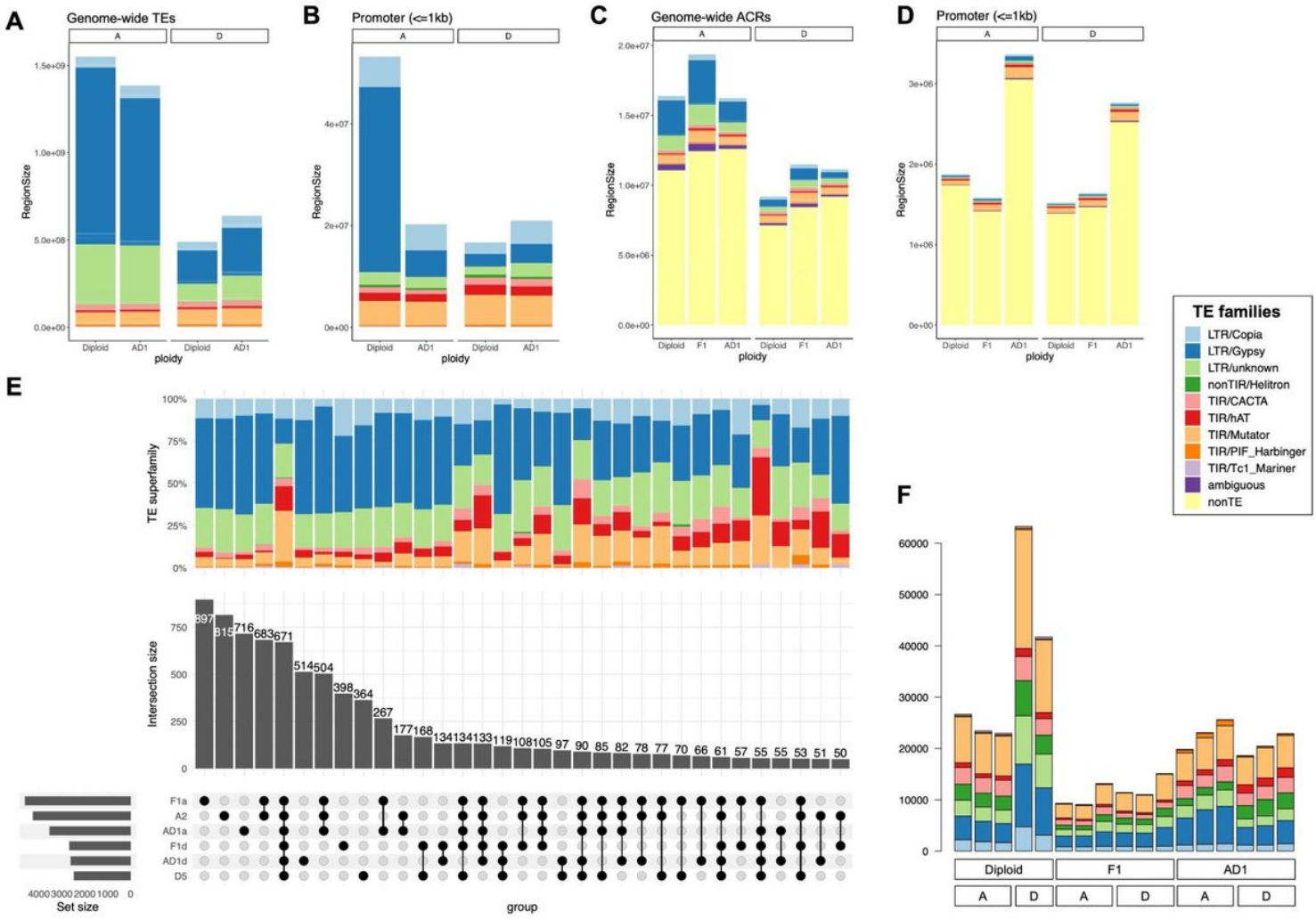


Figure 5

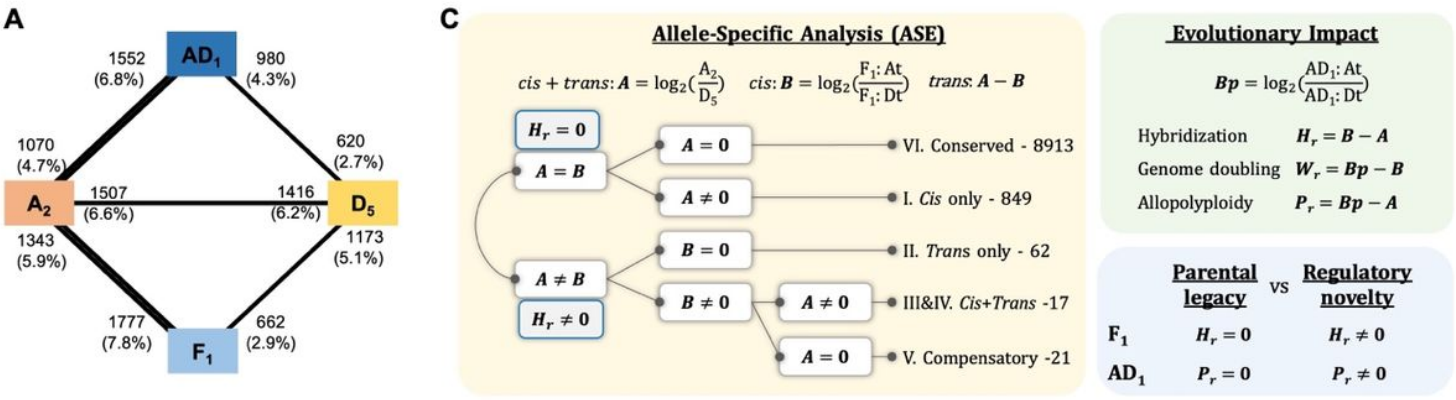
Chromatin accessibility and TEs.

A-B. Sizes of TE superfamilies at genome-wide scale (**A**) and within 1 kb promoters (**B**).

C-D. Contribution of TEs to ACRs at genome-wide scale (**A**) and within 1 kb promoters (**B**). ACRs that do not overlap with TEs were labeled as “nonTE”. ACRs that overlapped with TEs were considered TE-derived ACRs and further classified by TE superfamilies.

E. A union of 8,680 TE families significantly contributed to ACRs. A hypergeometric $P < 0.05$ of TE and ACR overlapping was required to consider the contribution of TE families. UpSetplot presents intersecting TE families sets, with the proportion of superfamilies shown in the barplot above.

F. TE expression (TPM) in cotton leaf transcriptomes.



B

	F_1	AD_1
A-dominant	537 (2.3%)	506 (2.2%)
D-dominant	1329 (5.8%)	1006 (4.4%)
Transgressive Up	429 (1.9%)	212 (0.9%)
Transgressive Down	202 (0.9%)	131 (0.6%)
Other nonadditive	738 (3.2%)	432 (1.9%)
Additive	13882 (60.6%)	11900 (52.0%)
Total	**17117 (74.8%)**	**14187 (62.0%)**

D

Phenomenon	Measure			Imbalance
		A>D	A<D	
A_2 vs D_5: Parental divergence	$A \neq 0$	1507 (6.6%)	1416 (6.2%)	-
F_1: Homoeolog expression bias	$B \neq 0$	807 (3.5%)	676 (3.0%)	A***
AD_1: Homoeolog expression bias	$Bp \neq 0$	1550 (6.8%)	1682 (7.3%)	D* (consistent)
		Up	Down	
Hybridization	$H_r \neq 0$	80 (0.3%)	62 (0.3%)	-
Genome Doubling	$W_r \neq 0$	283 (1.2%)	396 (1.7%)	Down***
Allopolyploidization	$P_r \neq 0$	527 (2.3%)	628 (2.4%)	Down**

Figure 6

Duplicated gene expression patterns based only on the consistent results by different mapping strategies.

A. Differential *total* expression of homoeologous genes in AD_1 and F_1 relative to A_2 and D_5 parental diploids. Between AD_1 and A_2, for example, 1552 genes (6.8% of 22,889 ortholog groups) are more highly expressed in AD_1, and 1070 genes (4.7%) are more highly expressed in A_2. The thicker lines relative to A_2 than to D_5 represented asymmetrically more expression changes to the A-genome parent.

B. Test of the additivity hypothesis in AD_1 and F_1. Non-additive expression categories include expression-level dominance (ELD; A-dominant and D-dominant), transgression (up and down regulation), and other non-additive patterns. For 74.8% and 62.0% of 22,889 ortholog groups, the classification results were consistent by different mapping strategies.

C. Illustration of the extended *cis* and *trans* analytic framework (Hu & Wendel, 2019), which combined the classic allele-specific analysis (ASE) with the estimation of evolutionary impact.

D. Summary of parental divergence, homoeolog expression bias (HEB), and evolutionary impacts. Chi-square tests were performed to infer the significance of imbalance: - as insignificant with $P > 0.05$; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Only for HEB in AD_1, the significant imbalance was found consistent by different mapping strategies, thus labeled as robust in parenthesis.

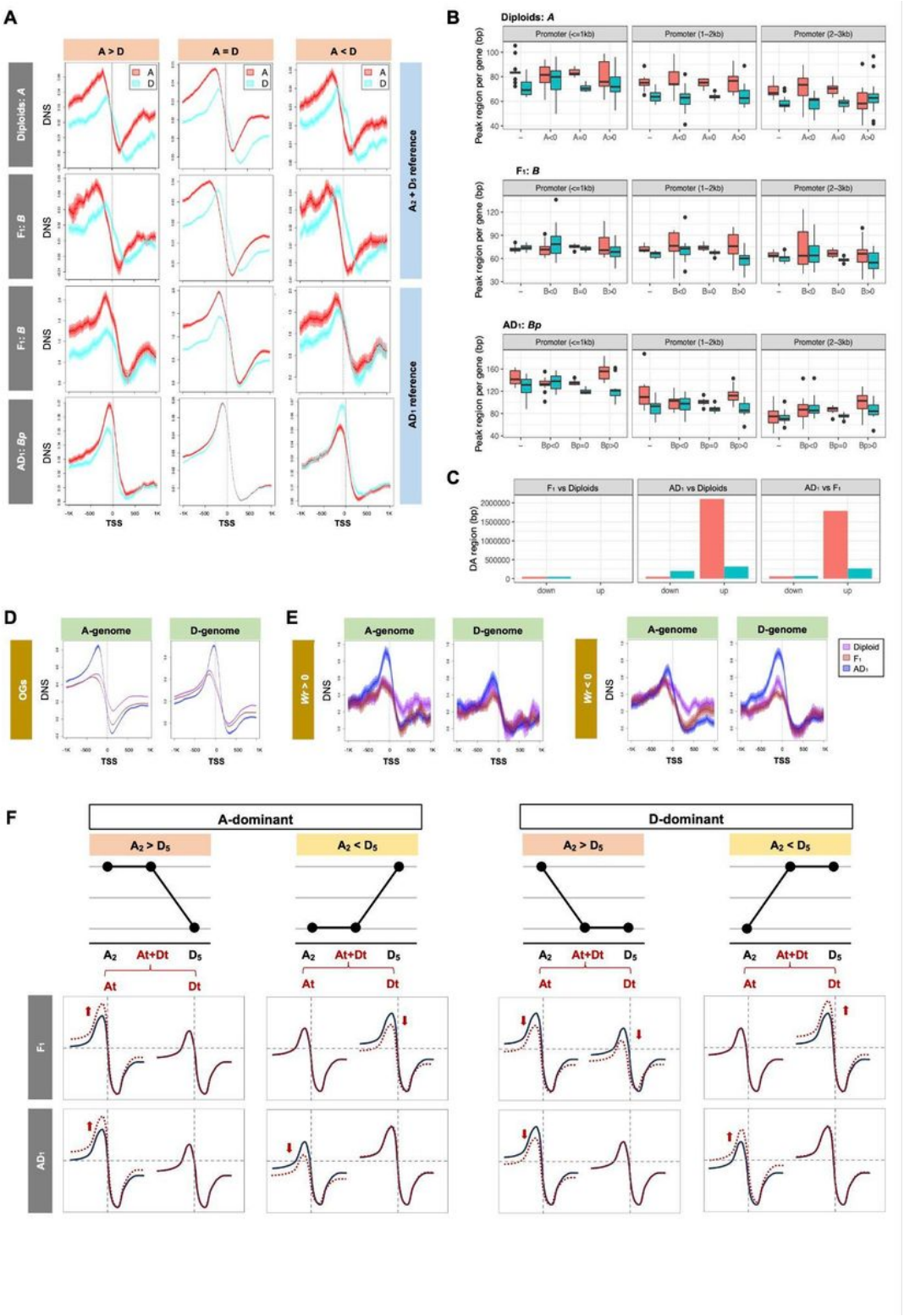


Figure 7

Promoter accessibility of duplicated genes in diploid and allopolyploid cottons.

A. Aggregation plots of DNS signals were present in association with duplicated gene expression patterns of parental divergence (***A*** ≠ 0; top row), homoeolog expression biases in F_1 (***B*** ≠ 0; middle two rows) and in AD_1 (***Bp*** ≠ 0; bottom row). The x-axis is centered on TSS ± 1 kb. The y-axis represents RPGC (reads per

genomic content) normalized occupancy performed by deepTools (Ramírez et al., 2014). Each center line represents the aggregated mean occupancy, with ribbons representing the 95% confidence interval.

B. Boxplot of promoter ACR sizes in association with duplicated expression patterns. Using the bottom row "AD_1: ***Bp***" as an example, promoter ACRs were further classified to three promoter regions (<=1 kb, 1-2 kb, and 2-3 kb) for presentation; within each panel, the ACR sizes per gene were contrasted between At and Dt for different expression patterns (***Bp*** > 0 indicates higher At versus Dt expression in AD_1, ***Bp*** < 0 indicates higher Dt versus At expression, and ***Bp*** = 0 indicates equal homoeolog expression; "-" refers to inconsistent results from different mapping strategies).

C. Bar plot of DA region sizes in pairwise comparisons between diploids, F_1, and AD_1 in 1 kb promoters. Within each plot panel, the increase and decrease of accessibility were plotted for A- (red) and D- (blue) genomes.

D. For 22,889 OGs, aggregation plots of DNS signals were presented based on A_2 and D_5 references.

E. For OGs exhibiting genome doubling effects on expression (283 ***Wr*** > 0 and 396 ***Wr*** < 0), aggregation plots of DNS signals were presented based on A- and D- subgenomes of AD_1 reference.

F. Corresponding to four ELD patterns, the modes of promoter accessibility changes were depicted for At and Dt homoeologs corresponding to their total expression patterns.

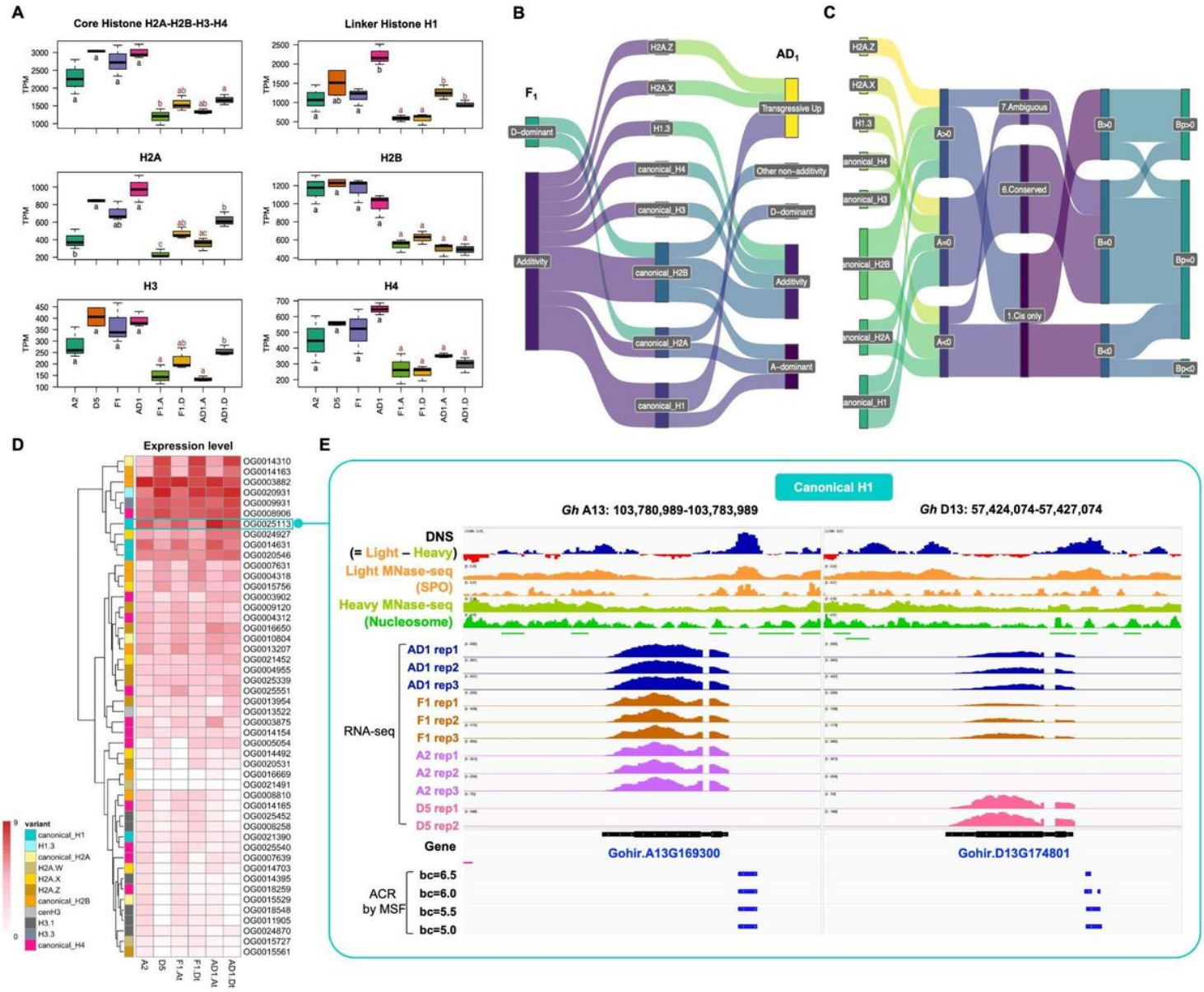


Figure 8

Analysis of histone gene expression.

A. Boxplots present summed expression levels of histone gene family. Comparisons across diploid and allopolyploid cottons. Comparisons between (sub)genomes were performed using ANOVA with post-hoc Tukey HSD test ($P < 0.05$). Groups with the same letter are not significantly different.

B. The inheritance mode of parental histone expression was compared between F_1 and AD_1, as characterized by additive and non-additive expression patterns (e.g. ELD and transgression).

Categorization of different histone variants for OGs was depicted by the middle level of the Sankey diagram.

C. Classifications of parental expression divergence (***A***), HEB in F_1 and AD_1 (***B*** and ***Bp***) were compared by Sankey diagram**.**

D. Heatmap of histone gene expression profiles of 47 OGs.

E. Genomic tracks illustrate dns-MNase-seq and RNA-seq profiles for a homoeologous pair of canonical H1 genes in *G. hirsutum.* Representatives of other histone variants were shown in Figure S13.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SuppleTables.xlsx.zip
- SupplementaryFiguresandTablesLegend.docx
- Supplementarytext.docx
- SuppleFigures.pdf
- Table1.docx