Evolutionary Dynamics of Chromatin Structure and Duplicate Gene Expression in Diploid and Allopolyploid Cotton

Guanjing Hu [1,2,*] Corrinne E. Grover [3] Daniel L. Vera,[4] Pei-Yau Lung,[5] Senthil B. Girimurugan,[6] Emma R. Miller,[3] Justin L. Conover [3,7,8] Shujun Ou,[9] Xianpeng Xiong,[2] De Zhu,[2] Dongming Li,[2,10] Joseph P. Gallagher,[11] Joshua A. Udall,[12] Xin Sui,[5] Jinfeng Zhang [5] Hank W. Bass [4,*] and Jonathan F. Wendel [3,*]

[1]State Key Laboratory of Cotton Bio-breeding and Integrated, Chinese Academy of Agricultural Sciences, Institute of Cotton Research, Anyang 455000, China

[2]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Chinese Academy of Agricultural Sciences, Agricultural Genomics Institute at Shenzhen, Shenzhen 518120, China

[3]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA

[4]Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

[5]Department of Statistics, Florida State University, Tallahassee, FL 32306, USA

[6]Department of Mathematics, Florida Gulf Coast University, Fort Myers, FL 33965, USA

[7]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

[8]Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA

[9]Department of Molecular Genetics, Ohio State University, Columbus, OH 43210, USA

[10]Zhengzhou Research Base, State Key Laboratory of Cotton Biology, School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450000, China

[11]Forage Seed and Cereal Research Unit, USDA/Agricultural Research Service, Corvallis, OR 97331, USA

[12]Crop Germplasm Research Unit, USDA/Agricultural Research Service, College Station, TX 77845, USA

***Corresponding authors:** Emails: jfw@iastate.edu; bass@bio.fsu.edu; huguanjing@caas.cn.

Associate editor: Michael Purugganan

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 intergenomic 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.

Key words: **allopolyploidy, chromatin accessibility, nucleosome organization, cotton, genome dominance, homoeolog expression bias.**

Received: November 02, 2023. **Revised:** April 10, 2024. **Accepted:** May 10, 2024

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 (Jiao et al. 2011; Ruprecht et al. 2017; Initiative & One Thousand Plant Transcriptomes Initiative 2019), whole-genome duplications resulting from polyploidy have significant implications for plant physiology, ecology, and evolution (Stebbins 1940; Levin 1983; Ramsey and Schemske 2002; Leitch and Leitch 2008; Wendel 2015; Soltis and Soltis 2016; Van de Peer et al. 2017, 2021; Wendel et al. 2018; Heslop-Harrison et al. 2022). Polyploidy may be associated with expanded ecological ranges (Arrigo et al. 2016; Coughlan et al. 2017; Baniaga et al. 2020; Wang et al. 2021; Parshuram et al. 2022; Zhao et al. 2022; Elliott et al. 2023; Mata et al. 2023), enhanced tolerance to biotic and abiotic stresses (reviewed in Van de Peer et al. 2021), physiological changes (Mishra 1997; Sugiyama 2005; Otto 2007; Knight and Beaulieu 2008; Coate et al. 2012; Orr-Weaver 2015), 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 (Olsen and Wendel 2013; Renny-Byfield and Wendel 2014; Zhang et al. 2019; Heslop-Harrison et al. 2022).

Increases in whole-genome content resulting from polyploidy are often associated with changes in nucleotypic characters, such as cell size, nuclear volume, and cell cycle duration (Wendel et al. 2018; Doyle and Coate 2019). 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 (Grover et al. 2012; Hu and Wendel 2019; Visger et al. 2019; Shan et al. 2020; Giraud et al. 2021). 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 (Zhang et al. 2023), apparent *trans*-control of duplicate expression (expression level dominance) (Rapp et al. 2009; Grover et al. 2012; Yoo and Wendel 2014; Yoo et al. 2014), and altered coexpression 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 (Talbert et al. 2019; Ahmad et al. 2022). 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 (Kornberg 1974; Andrews and Luger 2011). 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 (Zlatanova et al. 2008; Mieczkowski et al. 2016; Klemm et al. 2019). Thus, nucleosomes act as regulators of chromatin accessibility, which inherently manifests the myriad epigenetic modifications of histones and DNA that collectively control gene expression (Schmitz et al. 2011; Jordan and Schmitz 2016; Kawakatsu et al. 2016; Niederhuth et al. 2016; Hofmeister et al. 2017; Jackson 2017; Song et al. 2017; Springer and Schmitz 2017; Giles and Taberlay 2019; Klein and Hainer 2020). 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 (Tsompana and Buck 2014; Liu et al. 2015; Zhang et al. 2015; Voong et al. 2017; Zhang and Jiang 2018; Baldi et al. 2020; Galli et al. 2020; Jordan et al. 2020; Zhao et al. 2020; Barbier et al. 2021). 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 and Groudine 1976; Wu et al. 1979a, 1979b). High-throughput DHS mapping has provided genome-wide insight into *cis*-regulatory DNA elements (CREs) and TF binding sites (TFBSs) in various plant species (Zhang et al. 2012; Jiang 2015; Sullivan et al. 2015; Qiu et al. 2016; Zhao et al. 2018; Han et al. 2020, 2022). ATAC-seq, a more efficient alternative to DNase-seq, enables fast and low-input profiling of chromatin accessibility (Lu et al. 2017), 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; Ricci et al. 2019; Reynoso et al. 2022).

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 micrococcal nuclease (MNase) digest conditions, light and heavy, which provide 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 differential sensitivity MNase-seq (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 TFBSs. In maize, MNase hypersensitive regions account for <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 (Pass et al. 2017; Parvathaneni et al. 2020). 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*-regulatory elements within accessible chromatin regions (Savadel et al. 2021; Liang et al. 2022). Overall, the properties of MNase as a probe for chromatin structure have 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 (Wendel and Grover 2015; Hu et al. 2021; Viot and Wendel 2023), and new cotton species continue to be discovered (Stewart et al. 2015; Gallagher et al. 2017). Phylogenetic analyses (Wendel and Cronn 2003; Wendel et al. 2010; Chen et al. 2017) and genome sequence data (Huang et al. 2021) indicate that the genus originated ~5 to 10 million years ago (mya). 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 *Gossypium hirsutum* (Upland cotton) and *G. barbadense* (Sea Island cotton), each domesticated within the last 7,000 years (Wendel and Grover 2015). The closest extant species related to the D-genome progenitor is *Gossypium 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 (Rapp et al. 2009; Grover et al. 2012; Hu et al. 2013, 2014, 2015; Yoo et al. 2013; Gallagher et al. 2020). *Cis*- and *trans*-regulatory control of expression have also been studied in allopolyploid cotton, with *trans*-regulatory variants preferentially accumulating during about 5000 to 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 F_1 hybrid that mimics the natural hybridization that occurred 1 to 2 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.

Materials and Methods

Plant Materials

Four *Gossypium* genotypes were used, including a natural allopolyploid (AD genome), *G. hirsutum* cultivar Acala Maxxa (AD_1), and its model (A- and D-genome) diploid progenitors, i.e. *G. arboreum* accession A_2-101 (A_2) and *G. raimondii* (D_5). The two diploid genome groups, A and D, last shared a common ancestor 5 to 10 mya (Wendel and Albert 1992), and have diverged to the extent that genome sizes (GSs) differ 2-fold. Thus, the corresponding interspecific diploid F_1 hybrid ($A_2 \times D_5$) 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, IA, USA) under controlled short-day conditions (10 h photoperiod with darkness from 5 PM to 7 AM; 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 min (min) in 40 mL fixation buffer (1.0 M 2-methyl-2,4-pentanediol, 10 mM PIPES·NaOH at pH 7.0, 10 mM $MgCl_2$, 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 $MgCl_2$, 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 $MgCl_2$, and 1 mM $CaCl_2$) and distributed into 500 µL 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 (Yu et al. 2014), including *G. hirsutum* cv. TM1 UTX v2.1 (Chen et al. 2020), *G. arboreum* cv. SXY1 WHU-updated v1.0 (Huang et al. 2020) and *G. raimondii* JGI v2.0 (Paterson et al. 2012). The F_1 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 and 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 (TSSs) and transcription termination sites (TTSs). Read coverage data were 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) (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 and 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 nonoverlapped 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 NC and NRL values were calculated.

Mapping Accessible Chromatin Regions by DNS-seq

MNase Sensitive Footprints

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 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 (i) normalize the mapping read coverage in RPM between light and heavy MNase digestions, (ii) calculate DNS scores as the difference from light minus heavy read coverages, (iii) produce genome-browser-ready data tracks, and (iv) 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 (A_2 and D_5) and subgenomes (At and Dt) in hybrid and tetraploid cottons. A range of biological cutoff (BC) stringencies were tested in calling the MSFs and MNase 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, Supplementary Material online) to generate the final list of MSFs.

Subnucleosomal Particle Occupancy

As previously reported (Grossman et al. 2018; Savadel et al. 2021), small sequence fragments (0 to 130 bp) from the light MNase digestion can also be used to directly profile the occupancy of subnucleosome sized particles involved in transcriptional control. Using awk and BEDTools (v2.27.1) (Quinlan 2014), the geometric center of each small alignment (0 to 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 subnucleosomal particle occupancy (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, Supplementary Material online). The resulting list of segments represents accessible chromatin regions (ACRs) identified by SPO.

Mapping ACRs by ATAC-seq and DNase-seq

ATAC-seq

Two replicated ATAC-seq experiments were conducted using the young leaf tissue of *G. raimondii*, following a protocol described previously (Lu et al. 2017). For each replicate, approximately 200 mg freshly collected leaves or flash-frozen leaves were immediately chopped with a razor blade in 1 mL of prechilled 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. (2017). Sorted nuclei were incubated with 2 μL Tn5 transposase in a 40 μL tagmentation buffer (10 mM TAPS-NaOH pH 8.0, 5 mM $MgCl_2$) at 37 °C for 30 min without rotation. Integration products were purified using a Qiagen MinElute PCR Purification Kit or NEB Monarch DNA Clean-up Kit and then amplified using Phusion DNA polymerase for 10 to 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 (Wang et al. 2017, 2018; Han et al. 2022) and downloaded from NCBI (supplementary table S1, Supplementary Material online).

Data Processing

Raw ATAC-seq and DNase-seq reads were adapter and quality trimmed, and then filtered using "Trim Galore" (v0.4.5) (Krueger 2012). Clean reads were subsequently aligned to corresponding reference genomes using Bowtie2 (v2.3.4) (Langmead and 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, Supplementary Material online), 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) (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 ACRs for subsequent analyses.

Genrich Peak Calling

Postalignment 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, Supplementary Material online). 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 nonaccessible 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 to 2 kb, 2 to 3 kb), exon, intron, downstream (<1 kb, 1 to 2 kb, 2 to 3 kb), and intergenic regions (>3 kb upstream from TSS and >3 kb downstream from TTS).

Relative to Transposable Elements

Whole-genome transposable element (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 was 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 $\log_2$-transformed fold changes of observed versus the permutation-derived mean ACR proportions within TE superfamilies.

Differential Accessibility Analysis

Differences in chromatin accessibility attributable to hybridization and allopolyploidization were detected following an established differential accessibility (DA) workflow (Reske et al. 2020) using the R package *csaw* (v1.16.1) (Lun and Smyth 2016). For direct comparison between different cotton species, all MNase-seq data were aligned to the same reference genome, either the AD_1 reference genome or a concatenated reference of A_2 and D_5 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, Supplementary Material online). The resulting count matrices were then subject to the *edgeR* (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: (i) light versus heavy in diploids; (ii) light versus heavy in F_1; (iii) light versus heavy in AD_1; (iv) F_1:light−heavy versus diploids:light−heavy, representing hybridization effect; and (v) AD_1:light−heavy versus F_1: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 and Bailey 2021) and AME (McLeay and Bailey 2010). XSTREME conducts two types of de novo motif discovery using MEME and STREME followed by enrichment analysis using SEA (Bailey and 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 2019).

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 A_2, D_5, F_1, and AD_1 samples were generated with an average of 11 million read pairs per sample (supplementary table S1, Supplementary Material online). 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 an FDR $\alpha < 0.05$ required to identify significant changes.

To optimize the method to infer duplicated gene expression patterns, we tested the following mapping strategies. (i) *D_5-ref*: The *G. raimondii* (D_5) 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, D_5 reads were mapped to the D_5 transcripts; A_2 reads were mapped to the "pseudo-A_2" transcripts, which were generated by replacing species-diagnostic SNPs on the D_5 gene models with A_2-specific SNPs; F_1 reads were mapped against a concatenation of the pseudo-A_2 and D_5 transcripts; and *G. hirsutum* (AD_1) reads were mapped against a concatenation of pseudo-AD_1-A_t and pseudo-AD_1-D_t transcripts, which were similarly generated using AD_1-specific SNPs. (ii) *AD_1-ref*: reads from all species were individually mapped against the *G. hirsutum* (AD_1) transcript sequences (Chen et al. 2020). (iii) *individual-ref*: F_1 reads were mapped to the concatenated A_2 (Huang et al. 2020) and D_5 (Paterson et al. 2012) transcripts, while A_2, D_5, and AD_1 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. A_2, D_5, F_1:At, F_1:Dt, AD_1:At, and AD_1:Dt), which were inferred using the pSONIC pipeline (Conover et al. 2021) as previously described (Conover and Wendel 2022).

Based on the *total* (summed) expression of At and Dt homoeologs, F_1 and AD_1 gene expression was compared to expression in A_2 and D_5 and subsequently classified into following categories (Rapp et al. 2009): (i) additivity, whereby the *total* expression (in the hybrid or allopolyploid) is statistically equivalent to the mid-parent value of the parental diploids; (ii) A-genome ELD, whereby the *total* expression is statistically equivalent to the A_2 parent but different from the D_5 parent and mid-parent expression; (iii) D-genome ELD, whereby the *total* expression is statistically equivalent to the D_5 parent but different from the A_2 parent and mid-parent expression; (iv) transgressive up-regulation, whereby the *total* expression is greater than both A_2 and D_5; (v) transgressive down-regulation, whereby the *total* expression is less than both A_2 and D_5.

Based on the partitioned expression of At and Dt homoeologs (separately), HEB was assessed in the F_1 and AD_1 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 $\log_2$ ratios of A_2 and D_5 [$\boldsymbol{A} = \log_2(A_2/D_5)$], the *cis* effects by $\log_2$ ratios of their corresponding homoeologs [$\boldsymbol{B} = \log_2(\mathrm{At}/\mathrm{Dt})$], 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 Fig. 6c. The evolutionary impact of hybridization (***Hr***), allopolyploidization (***Pr***), and genome doubling (***Wr***) was determined according to Hu and Wendel (2019) and as illustrated in Fig. 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. (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 1,000 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

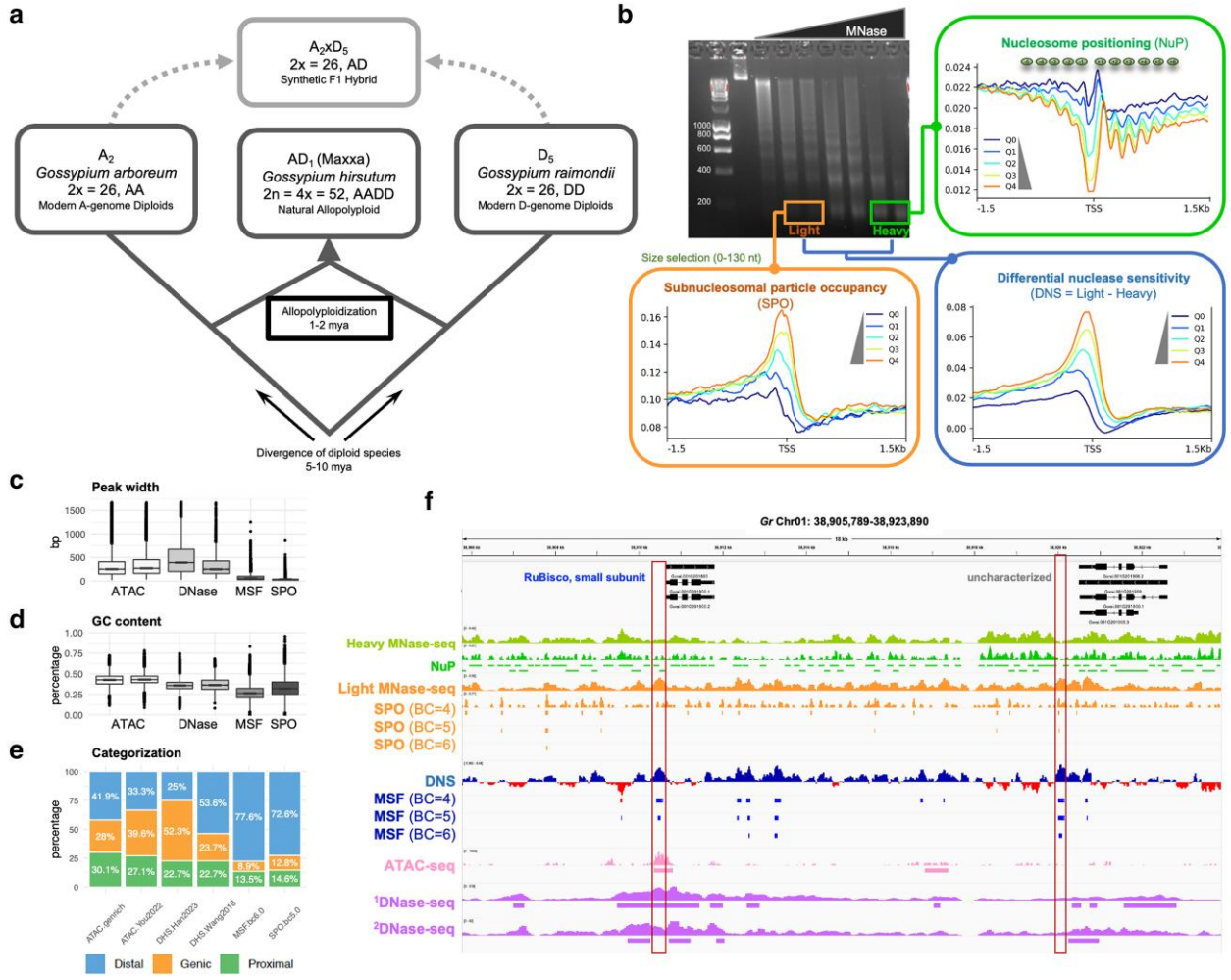


Fig. 1. Studying chromatin structure evolution in diploid and allopolyploid cottons. a) Four *Gossypium* genotypes were used in this study: a natural allopolyploid, *G. hirsutum* cultivar Acala Maxxa (AD_1); the model A- and D-genome diploid progenitors—*G. arboreum* accession A_2-101 (A_2) cultivar and *G. raimondii* (D_5); and their corresponding interspecific diploid F_1 hybrid ($A_2 \times D_5$). b) The technique of DNS-seq was used to profile various chromatin features, including nucleosome positioning (NuP), SPO, and 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 TSS and binned by five gene expression level groups, where Q1 to Q4 represent increasing expression quantiles, and Q0 represents the group of nonexpressed genes. c to e). ACRs were compared between the analyses of MSF, SPO, ATAC-seq, and DNase-seq (see supplementary table S6, Supplementary Material online), 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 shows 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 boxes.

PRJNA529417. All data used are detailed in supplementary table S1, Supplementary Material online. Custom scripts are available at the following GitHub repository: https://wendellab.github.io/cottonMNase-seq/.

Results

Mapping Chromatin Landscapes by Differential Sensitivity MNase-seq

To characterize the genome-wide chromatin features and *cis*-regulatory landscapes, we performed 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 (AD_1; GS = 2.2 Gb), A-genome diploid *G. arboreum* accession A_2-101 (A_2; GS = 1.8 Gb), D-genome diploid *G. raimondii* (D_5; GS = 0.8 Gb), and their synthetic F_1 hybrid ($A_2 \times D_5$; 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 GS per library, resulting in 591 million A_2, 126 million D_5, 549 million $A_2 \times D_5$, and 685 million AD_1 read pairs (supplementary table S2, Supplementary Material online). After adapter trimming and quality

filtering, the remaining 91% to 98% of reads were mapped to their corresponding reference genomes. Interestingly, the proportion of high-quality alignments ($Q > 20$) was notably higher for the D_5 reads (80% to 86%) versus the other genomes surveyed (range: 60% to 74%; supplementary table S2, Supplementary Material online), likely reflecting lower repetitive content of the smaller D_5 genome. Quality evaluation of the mapping results indicates that genomic coverage profiles were highly correlated between biological replicates ($R^2 = 0.91$ to 0.99); therefore, alignments from replicates (per species and per digestive condition) were combined in the following analyses.

As illustrated in Fig. 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% to 20% of each *Gossypium* genome (supplementary table S3, Supplementary Material online), 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% to 70% of each cotton genome, likely reflect positional variability and dynamics in multicellular samples. Around the TSS, the canonical pattern of nucleosome occupancy was observed: (i) 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; (ii) 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; (iii) highly expressed genes tend to have a lower degree of nucleosome occupancy and a larger NFR (Fig. 1b).

The light MNase digestion releases more sensitive, "fragile" nucleosomes and subnucleosomal sized particles (e.g. TFs), which have been used to map MNase hypersensitive sites (MHSs) and profile chromatin accessibility as a complementary approach to DNase-seq and ATAC-seq (Pass et al. 2017; Parvathaneni et al. 2020; Zhao et al. 2020; Savadel et al. 2021). Here, the smaller DNA fragments (0 to 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 SPO (Fig. 1b, lower left), as per Teves and Henikoff (2011), these regions are sometimes referred to as "MHSs" (Zhao et al. 2020) or "MFs" for MOA-seq footprint (Savadel et al. 2021) regions. The DNS (Fig. 1b, lower right) approach permits the identification of MSFs that reveal *cis*-regulatory landscapes (Rodgers-Melnick et al. 2016; Parvathaneni et al. 2020). 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 each *Gossypium* genotype studied (supplementary fig. S1, Supplementary Material online).

Comparing Accessibility Analysis by DNS-seq with ATAC-seq and DNase-seq

To assess chromatin accessibility profiles obtained by DNS-seq, we compared our MSF and SPO results with independent datasets generated through different enzymatic assays, ATAC-seq and DNase-seq. Our comprehensive comparison included two replicated ATAC-seq experiments performed in this study (supplementary text 2, Supplementary Material online) and integrated publicly available ATAC-seq and DNase-seq data from three independent studies (Wang et al. 2018; Han et al. 2022; You et al. 2022), all focusing on young leaves of *G. raimondii* (supplementary table S1, Supplementary Material online). Applying recommended analyses of quality control metrics (Bubb and Deal 2020; Schmitz et al. 2022), we observed substantial variations across these datasets in sequencing depth (16.2 to 250.2 million read pairs per library), mapping rate (88.2% to 95.5%), duplication read rate (6.1% to 87.2%), and signal-to-background ratio (18.4% to 45.6%) (supplementary table S4, Supplementary Material online). The number of identified ACRs ranged from 2,059 to 59,763 (supplementary table S5, Supplementary Material online), with overlaps between datasets falling between 3% and 96% (supplementary table S6, Supplementary Material online). Such extensive variations highlight the potential influence of experimental methods and inherent challenges in chromatin accessibility assays for cotton, likely due to its high polyphenol and polysaccharide content. It is important to consider these factors when next comparing results from different techniques.

To explore the relationships between accessibility profiles mapped by different assays, we performed heatmap clustering of Pearson correlation coefficients and principal component analysis (PCA), which revealed distinct clusters for the genome-wide maps generated by DNS-seq, separated from those by ATAC-seq and DNase-seq (supplementary fig. S3, Supplementary Material online). Additionally, only a small fraction (<11%) of MSF and SPO regions were covered by ATAC-seq or DNase-seq peaks (supplementary table S6, Supplementary Material online), suggesting unique features captured by DNS-seq. Furthermore, the ACRs detected by DNS-seq exhibited several distinct characteristics compared to those identified by ATAC-seq and DNase-seq. Notably, MSF and SPO peaks were smaller, had lower GC content, and exhibited a more prominent distribution distal to genes (Fig. 1c to e; supplementary table S5, Supplementary Material online). These smaller ACRs with sharper signals are typically preferred for identifying *cis*-regulatory motifs (Savadel et al. 2021), and lower GC content has been shown to be more indicative of accessible *cis*-regulatory regions flanking genes (Gaffney et al. 2012; Ando et al. 2019; Weinberg-Shukron et al. 2022). Importantly, the accessibility profiles by MSF and SPO demonstrated the expected enrichment before the TSSs and depletion the gene bodies, showing a positive correlation with

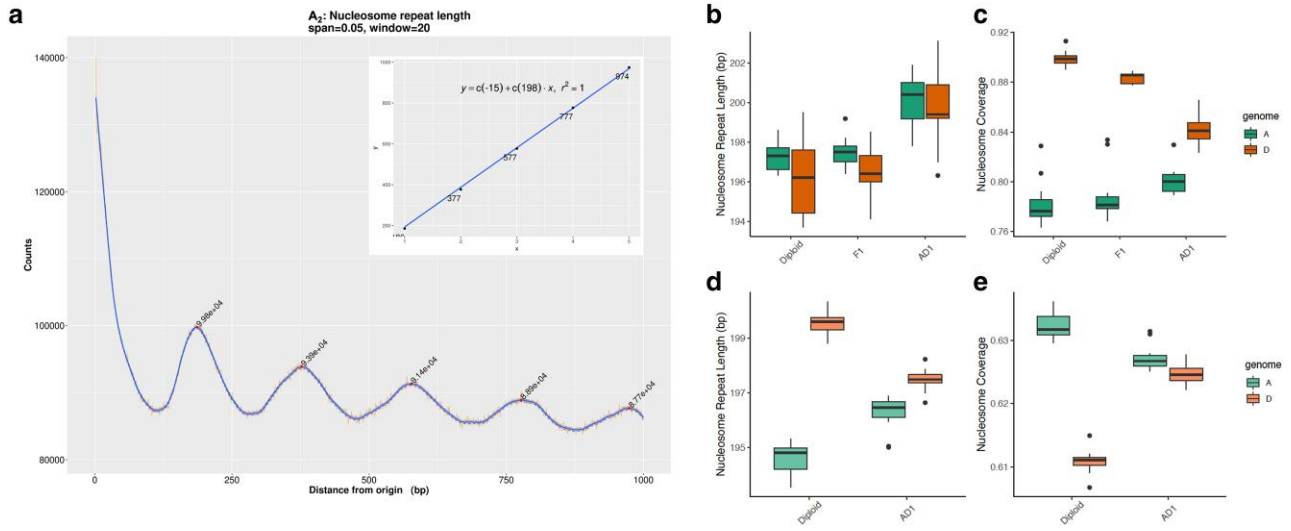


Fig. 2. Comparing nucleosome organization in diploid, hybrid, and allopolyploid cottons. a) Nucleosome phasogram exhibits a wave-like pattern of distances between neighboring nucleosome centers. Inset presents a linear fit to the positions of the phase peaks, where the slope represents the estimated NRL of 198 bp in the exemplar, A_2. b) Estimated NRL by phasogram across diploid and polyploid cotton genomes. c) Estimated NC based on the nucleosome positioning profiled by MNase-seq under heavy digestive conditions. d and e) Predicted NRL and NC based on reference genome sequence, respectively.

gene expression levels (supplementary fig. S1, Supplementary Material online), consistent with previous findings (Buenrostro et al. 2013). In contrast, the ATAC-seq and DNase-seq datasets often exhibited enrichment within gene bodies rather than before TSSs and lacked less robust correlations with gene expression (supplementary fig. S2, Supplementary Material online). These results suggest that DNS-seq offers a valuable approach for mapping chromatin accessibility with a strong link to gene expression levels in cotton.

In summary, the distinct clustering of genome-wide profiles (supplementary fig. S3, Supplementary Material online) and limited overlaps in ACRs (supplementary table S6, Supplementary Material online) highlight the unique perspectives provided by DNS-seq profiles, particularly in distant nongenic regions that appear to be less well represented in assays based on Tn5 or DNase I (Fig. 1e). This observation aligns with the previous studies of MHSs (equivalent to the SPO footprints here) in *Arabidopsis* (Zhao et al. 2020) and soybean (Fang et al. 2023), where a significant portion of MHSs (22% in *Arabidopsis* and 67% to 77% in soybean) were not detected by ATAC-seq or DNase-seq. The higher specificity observed here (89% of MSF and SPO peaks unique to DNS-seq) might be partially explained by the technical challenges and data quality issues associated with ATAC-seq and DNase-seq in cotton, as evidenced by the data quality variation (supplementary table S4, Supplementary Material online) and atypical ACR enrichment within gene bodies (supplementary fig. S2, Supplementary Material online). More importantly, these findings suggest that MNase-based approaches like DNS-seq hold significant promise for confident profiling of chromatin accessibility, particularly in resistant plants like cotton. This advantage likely stems from the utility of different MNase digestion conditions to derive reliable estimates of chromatin accessibility, in addition to the ability to capture MNase-specific sites in distal nongenic regulatory regions.

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 NRL, was estimated by applying a linear model to calculate the phasogram period (Fig. 2a; supplementary fig. S4 and table S7, Supplementary Material online). 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 (F_1) versus the allopolyploid cotton (AD_1) and that the D-genome NRLs were generally shorter than those in the A-genome (Fig. 2b; Diploids: A_2 197.3 ± 0.2 bp, D_5 196.2 ± 0.5 bp; F_1: At 197.5 ± 0.2 bp, Dt 196.4 ± 0.4 bp; AD_1: At 200.1 ± 0.4 bp, Dt 199.7 ± 0.5 bp; ANOVA followed by Tukey's post hoc test, $P < 0.05$: diploids = F_1 < AD_1 and D < A). Consistent with these observations, the percentage of genomic regions occupied by nucleosomes (i.e. NC) also exhibited

lower A- versus D-coverage, regardless of ploidy (Fig. 2c; Diploids: A_2 78.2 ± 0.5%, D_5 89.9 ± 0.1%; F_1: At 78.8 ± 0.6%, Dt 90.0 ± 0.21%; AD_1: 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 A_2 and D_5 diploids relative to the inter-subgenomic differences between At and Dt in the allopolyploid (AD_1), 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 and 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. A_2, D_5, and AD_1) suggest that the NRL should be shorter in the A-(sub)genomes (vs. the D-(sub)genomes) with a concomitantly higher NC value, regardless of ploidy level (Fig. 2d and e). This observation directly contrasts the experimentally observed pattern (Fig. 2b and c) 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 A_2 and D_5 diploids exhibit significant reductions in the At and Dt subgenomes of the allopolyploid (AD_1). 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% to 1.4% of each genome (Table 1; supplementary tables S8 to S11, Supplementary Material online). In the F_1 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, A_2 (296,312; 16.4 Mbp) and D_5 (190,795; 9.2 Mbp). In the allopolyploid AD_1, 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% to 87%), whereas 10% to 19% occurred proximally within the 2 kb gene flanking regions (proximal, pACRs) and only 4% to 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 (Fig. 3a), consistent with observations in other plant species which suggest that the proportion of dACRs is positively correlated with GS (Lu et al. 2019). This correlation with GS was even more significant for the total length of dACRs (Fig. 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 were significantly increased in AD_1, specifically due to expansions in the 1 kb promoter regions (Fig. 3c and d).

For each genome, an initial scan of the 1 kb promoter sequences for known DNA motifs from plantTFDB v5.0 (Jin et al. 2017) revealed relatively consistent motif occurrences across (sub)genomes, although the A_2 promoters exhibited the most divergence relative to the other genomes (Fig. 4a), possibly due to the elevated GC content in its promoters (A_2 30.56%, vs. AD1:At 28.08%, D_5 28.72%, AD_1: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 F_1, and in AD_1, respectively (supplementary table S12, Supplementary Material online). 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 (Fig. 4b). Interestingly, AD_1-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 TFBSs of MYB (ten motifs), WRKY (nine motifs), bZIP (nine motifs), and TCP (nine motifs) TF families (Fig. 4c; supplementary table S12, Supplementary Material online). Congruently, a heatmap dendrogram of pACR motif enrichment rankings showed that AD_1:At and AD_1:Dt were more similar to each other and distinct from the diploid enrichment rankings. Among the diploids, clustering of F_1:Dt and D_5 showed their higher similarity, with the A_2 and F_1:At genomes falling more basally in that clade (Fig. 4b). Furthermore, de novo motif discovery by XSTREME and clustering analysis (supplementary fig. S5, Supplementary Material online) 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 TEs revealed that the A subgenome of AD_1 has 1.2% lower TEs than A_2 whereas the

Table 1 ACR classification

	Diploid		F_1		AD_1	
	A_2	D_5	At	Dt	At	Dt
Number	296,312	190,795	357,171	224,483	260,278	189,068
In proximity to genes						
gACR	19,025 (6.4%)	22,240 (11.7%)	13,948 (3.9%)	17,754 (7.9%)	18,540 (7.1%)	17,151 (9.1%)
pACR	30,710 (10.4%)	26,881 (14.1%)	33,873 (9.5%)	29,719 (13.2%)	38,907 (14.9%)	35,333 (18.7%)
dACR	246,577 (83.2%)	141,674 (74.3%)	309,350 (86.6%)	177,010 (78.9%)	20,2831 (77.9%)	136,584 (72.2%)
Overlapped with TEs						
LTR retrotransposons	75,406 (25.4%)	24,306 (12.7%)	98,680 (27.6%)	36,251 (16.1%)	48,132 (18.5%)	20,486 (10.8%)
DNA transposons	19,161 (6.5%)	17,526 (9.2%)	26,249 (7.3%)	23.941 (10.7%)	16,658 (6.4%)	15,738 (8.3%)
Nonoverlapped with TEs	201,745 (68.1%)	148,963 (78.1%)	232,242 (65.0%)	164,291 (73.2%)	195,488 (75.1%)	152,844 (80.8%)
Length (Mbp)	16.4	9.2	19.3	11.5	16.2	11.1
In proximity to genes						
gACR	1.0 (6.3%)	1.0 (11.0%)	0.7 (3.6%)	0.9 (7.8%)	1.2 (7.6%)	1.0 (9.3%)
pACR	1.7 (10.6%)	1.3 (14.3%)	1.8 (9.5%)	1.6 (13.6%)	2.7 (16.8%)	2.3 (20.4%)
dACR	13.6 (83.1%)	6.9 (74.7%)	16.8 (86.9%)	9.0 (78.6%)	12.3 (75.6%)	7.8 (70.3%)
Overlapped with TEs						
LTR retrotransposons	4.2 (25.8%)	1.2 (13.0%)	5.4 (28.1%)	1.8 (16.0%)	2.7 (16.3%)	1.1 (9.8%)
DNA transposons	1.1 (6.7%)	0.9 (9.4%)	1.5 (7.8%)	1.2 (10.8%)	1.0 (6.2%)	0.9 (7.9%)
Nonoverlapped with TEs	11.1 (67.5%)	7.1 (77.6%)	12.4 (64.2%)	8.4 (73.2%)	12.6 (77.5%)	9.2 (82.4%)

D subgenome has 5.1% more TEs than D_5 (AD_1:At = 81.2%, AD_1:Dt = 64.6%, A_2 = 82.4%, and D_5 = 59.5%; Fig. 5a; supplementary table S13, Supplementary Material online), consistent with previous reports (Zhao et al. 1998; Chen et al. 2020). ACRs accounted for only 0.31% to 0.68% of genomic regions annotated as TEs, significantly lower than their composition in other genomic regions (1.09% to 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 (Fig. 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 (AD_1) contained the lowest amounts of ACRs that overlapped with TEs (AD_1: At 22.5%, Dt 17.6%; F_1: At 35.8%, Dt 26.8%; A_2 32.5%; D_5 22.4%), indicating decreased chromatin accessibility in TE regions accompanying allopolyploidization.

Because the allopolyploid (AD_1) exhibits both a reduction in TE-overlapping ACRs and an increase in promoter ACRs (Fig. 3d), we hypothesized that promoters may have gained more accessibility from TE removal associated with polyploidization. The general distribution of TEs around TSSs is similar between diploid and polyploid cottons (supplementary fig. S6, Supplementary Material online). However, the diploid A_2 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 (Fig. 5b; supplementary fig. S7 and table S14, Supplementary Material online). At the genome-wide scale, TEs contributed to 18% to 36% of ACRs (Table 1; Fig. 5c), but these accessible TEs were mainly located in distal intergenic regions and only contributed to a small portion of promoter ACRs (Fig. 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 AD_1 promoters (Fig. 5b) is associated with a gain of both non-TE and TE-derived ACRs accompanying allopolyploidy (Fig. 5d). The percentage of TE-derived ACRs at the 1 kb promoter regions was significantly higher in AD_1 than in diploids and F_1 (AD_1: At 0.86%, Dt 0.63%; F_1: At 0.41%, Dt 0.47%; A_2 0.46%; D_5 0.36%; χ^2 test $P < 0.05$). Although this observation supports our hypothesis that promoter TE depletion led to increased accessibility in the At genome of the allopolyploid (relative to A_2), it does not explain the increased accessibility in the Dt genome (vs. D_5).

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 (AD_1:At 0.96, AD_1:Dt 0.84; F_1:At 0.98, F_1:Dt 0.96; A_2 0.99; and D_5 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 was found significantly enriched in TE-derived ACRs (Fig. 5e). Intersection of TE families among genomes revealed a significant proportion of lineage-specific TE families, which accounts for 14% to 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% to 6% of mapped RNA-seq reads. Notably,

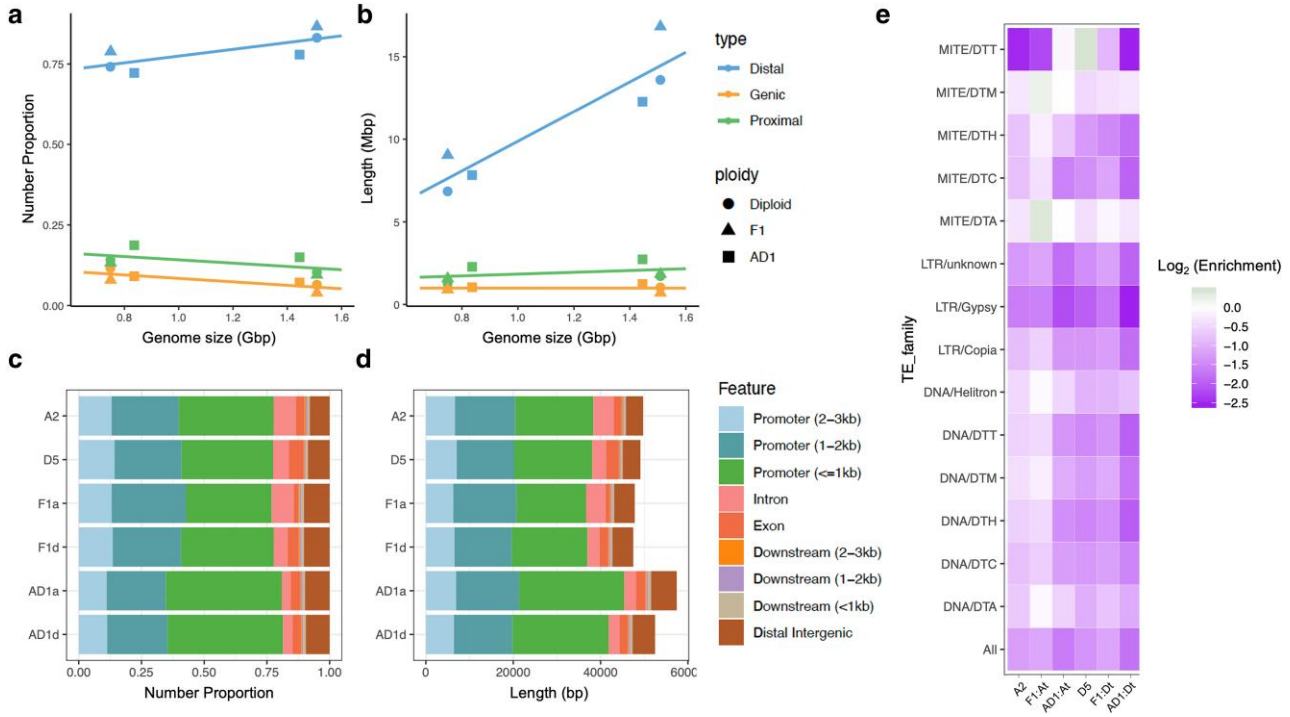


Fig. 3. Comparing ACRs in diploid, hybrid, and allopolyploid cotton. a and b) ACRs were categorized as genic (gACRs), proximal (pACRs), or distal ACRs (dACRs). Their relative proportions a) and total lengths b) are presented (*y*-axis) against corresponding genome sizes (*x*-axis), with a linear regression 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 and 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 1,000 permutations) mean ACR proportions within TE superfamilies.

4,622 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 (χ^2 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 (A_2 3622, D_5 2973, F_1:At 3242, F_1:Dt 3115, AD_1:At 3381, AD_1:Dt 2946), higher transcript abundances were found in D_5 (Fig. 5f; supplementary fig. S8, Supplementary Material online).

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 A_2 and D_5 genomes (supplementary tables S8 and S9, Supplementary Material online) with their homologs in the At and Dt subgenomes of the interspecific diploid hybrid (F_1) (supplementary table S11, Supplementary Material online). This comparison resulted in only 10.5% of diploid ACRs (11.0% of A_2 and 9.7% of D_5) overlapping with F_1 ACRs. We note that the relatively low overlap in ACRs between diploids and F_1 reflects the stringency of our peak calling method. Notably, as stringency increased, overlaps decreased, as detailed in supplementary text 1, Supplementary Material online. Given the stringent criteria applied independently for each genotype, the outcome of limited overlap was not entirely unexpected from this conservative approach.

To further validate our findings, we reanalyzed the DNase-seq diploid and F_1 datasets reported by Han et al. (2022) for comparison (supplementary table S1, Supplementary Material online). This analysis detected 28,029 (8.7 Mbp), 59,763 (34.1 Mbp), and 36,467 (11.2 Mbp) differential hypersensitive sites (DHSs) from A_2, D_5, and F_1, respectively. In comparison with the original report (77,915 in A_2, 59,997 in D_5, and 78,722 in F_1), our analysis is over 2-fold more stringent, resulting in much lower overlap of 36.6% between diploids and F_1 compared to the reported overlap of 83.7%. Additionally, much smaller ACRs were detected by DNS-seq compared to those identified by DNase-seq (mean peak width 54 bp vs. 492 bp), which naturally leads to reduced overlap upon intersection.

Recognizing these technical nuances, we employed a DA approach directly contrasting the MNase-seq data between diploid genomes and their corresponding subgenomes in the F_1 and natural allopolyploid (see Materials and Methods and supplementary text 4, Supplementary Material online). Compared to ACR intersections, this method is more sensitive to capturing similarities between accessibility profiles and independent of the stringency of ACR peak calling. The DA analysis of allopolyploid versus

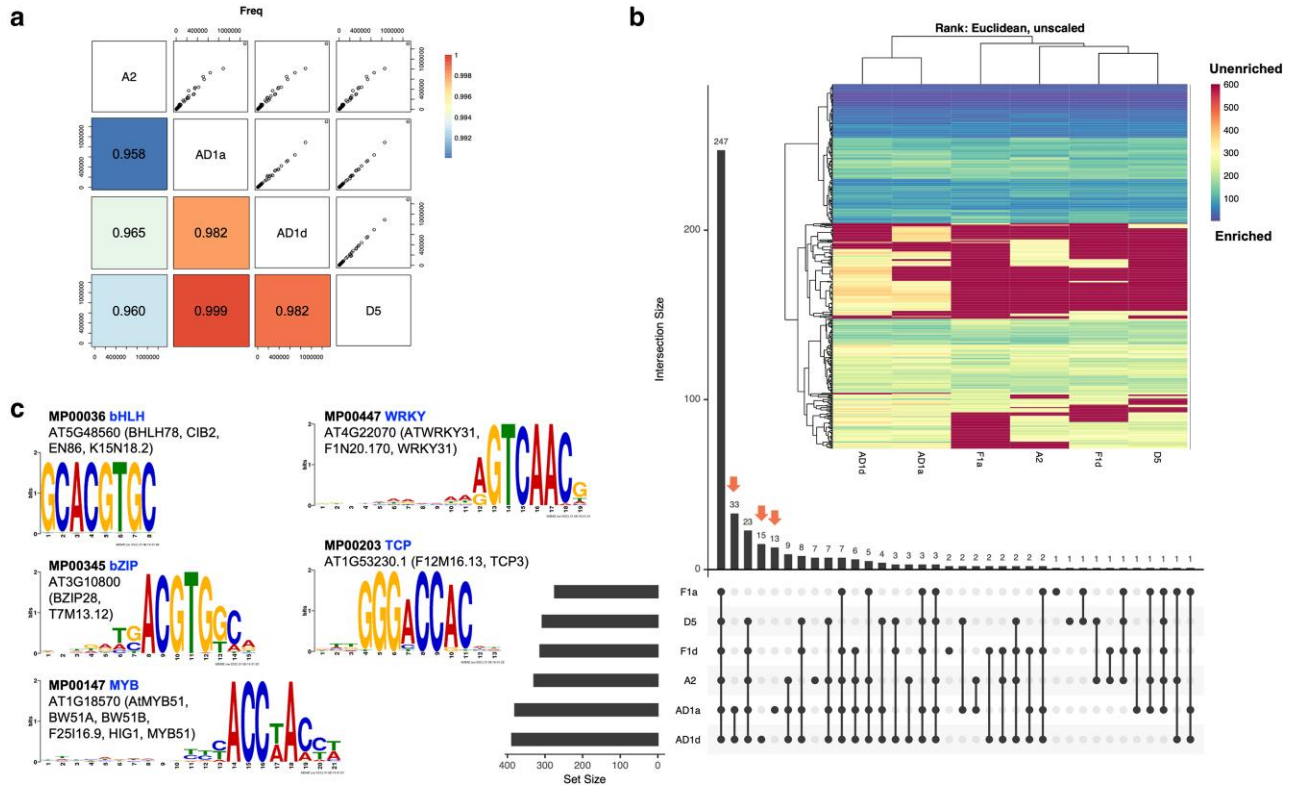


Fig. 4. Motif analysis of pACRs in 1 kb promoters. a) Comparing background promoter sequences between genomes, based on FIMO scanning of their whole promoter region sequences for known motifs from plantTFDB v5.0 (Jin et al. 2017). Upper triangle of the matrix shows the scatter plots of motif frequency, and the lower triangle presents the pairwise Pearson correlation coefficients of motif frequency. b) A union of 423 enriched motifs within the 1 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 was assigned to unenriched motifs in the corresponding genome. c) Top 5 most enriched AD_1-specific motifs.

diploids revealed an increase of 3.3 to 4.4 Mb and a decrease of 435 to 740 kb in accessibility. In contrast, the differences between the F_1 hybrid and the diploids were smaller, showing an increase of 16 to 304 kb and a decrease of 132 to 214 kb. These results indicate that allopolyploidization and subsequent evolution at the allopolyploid level, for >1 to 2 million years in this case, collectively induce much greater changes in chromatin accessibility than does hybridization in the F_1. 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 A_2, D_5, their F_1 hybrid, and natural allopolyploid derivative, AD_1 (supplementary table S1, Supplementary Material online). Duplicated gene expression patterns (Box 1) were categorized under a preestablished analytical framework (Hu and Wendel 2019), illustrated in Fig. 6. We employed a conservative approach and only report results that are consistent across different mapping strategies (supplementary tables S15 to S18 and text S5, Supplementary Material online). 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 A_2, D_5, F_1:At, F1:Dt, AD_1:At, and AD_1:Dt), duplicated gene expression patterns were characterized based on *total* (Fig. 6a and b) and partitioned homoeologous expression levels (Fig. 6c and d), including differential *total* expression relative to parental diploids, ELD, HEB, *cis*- and *trans*-regulatory divergence, as well as the evolutionary impact of hybridization (***Hr***), allopolyploidization (***Pr***), and genome doubling (***Wr***).

In both the F_1 and AD_1, the *total* expression of homoeologous genes exhibited more differential expression relative to A_2 than to D_5 (F_1—13.7% vs. 8.0%; AD_1—11.5% vs. 7.0%; Fig. 6a), and, correspondingly, the ELD analysis revealed more D-dominant than A-dominant expression patterns (F_1—5.8% vs. 2.3%; AD_1—4.4% vs. 2.2%; Fig. 6b). These observations suggest an asymmetric resemblance of the overall transcriptome toward the D-genome diploid parent, as noted previously (Flagel et al. 2008; Rapp et al. 2009; Yoo

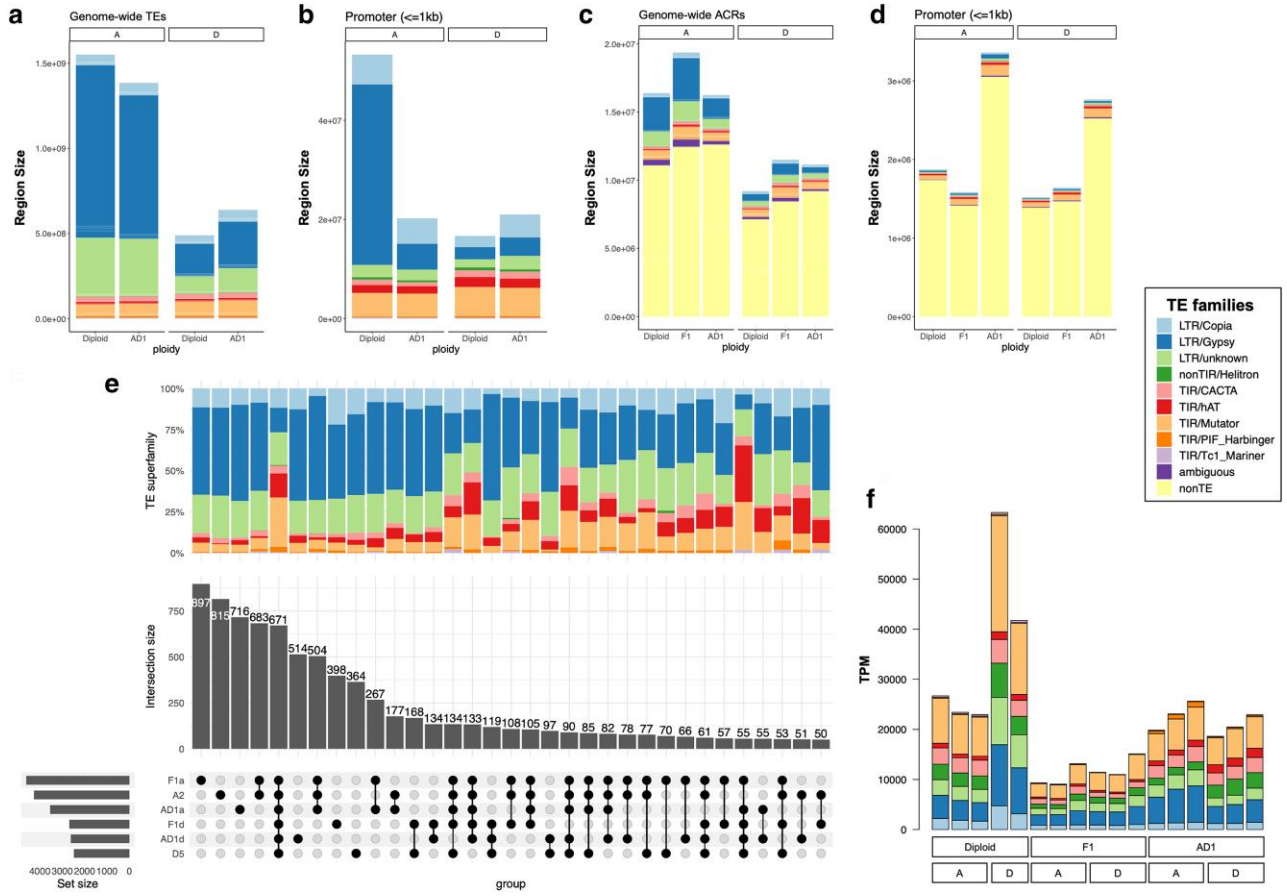


Fig. 5. Chromatin accessibility and TEs. a and b) Sizes of TE superfamilies at genome-wide scale a) and within 1 kb promoters b). c and d) Contribution of TEs to ACRs at genome-wide scale a) and within 1 kb promoters b). ACRs that do not overlap with TEs were labeled as "non-TE." ACRs that overlapped 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 measured in TPM reads using the cotton leaf transcriptome data.

and Wendel 2014). This trend was consistent across different mapping strategies (supplementary tables S16 to S18, Supplementary Material online).

When expression was compared between homoeologs, HEB was detected for 6.5% ($\boldsymbol{B} \neq 0$) and 14.1% ($\boldsymbol{Bp} \neq 0$) of genes in the F_1 and AD_1, respectively (Fig. 6d; supplementary table S18, Supplementary Material online), representing a greater than 2-fold increase in HEB in the allopolyploid. While no overall imbalance in HEB was detected in the F_1, in the allopolyploid more homoeologous pairs exhibited D- (vs. A-) biases, regardless of mapping strategy. Allele-specific expression (ASE) analysis revealed 12.8% of genes exhibited parental expression divergence between A_2 and D_5, whose inferred regulation can be subdivided into the previously described categories (Hu and 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 (supplementary table S18, Supplementary Material online), 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 ($\boldsymbol{Wr} \neq 0$, 2.9%) has a much stronger effect than hybridization ($\boldsymbol{Hr} \neq 0$, 0.6%), representing two distinct phases of allopolyploidization ($\boldsymbol{Pr} \neq 0$, 4.7%). These results also suggested that the relative expression of At versus Dt homoeologs in F_1 and AD_1 was mainly determined by the parental state of A_2 versus D_5 ($\boldsymbol{Hr} = 0$, 99.4%; $\boldsymbol{Pr} = 0$, 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 ($\boldsymbol{Hr} \neq 0$ and $\boldsymbol{Pr} \neq 0$, as illustrated in Fig. 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), the parental A_2 and D_5 accessibility profiles were generated by mapping diploid DNS-seq reads to their corresponding

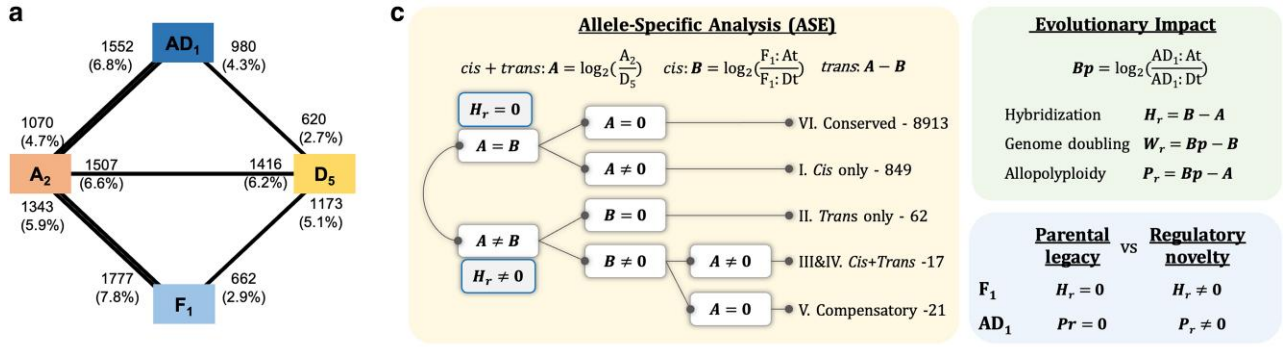
Box 1. Terminology of duplicated gene expression patterns

- Additive and nonadditive expression: A condition of allopolyploid or hybrid gene expression relative to parental expression levels. Additive expression is inferred when the *total* expression of homoeologous copies matches the arithmetic average of orthologous expression levels in parental diploids. Nonadditive expression deviates from this average, characterizing novel expression patterns specific to a hybrid or allopolyploid.
- ELD: A subcategory of nonadditive expression, where the *total* homoeologous expression is equivalent to the orthologous expression level of one parental diploid but not the other; in the present application, when the aggregate (At + Dt) expression level in the allopolyploid is statistically equivalent to the level of the A genome but not the D-genome diploid, the allopolyploid exhibits ELD toward the A-genome, or A-dominant expression. Note that ELD can be inferred for both higher and lower expression levels of one diploid relative to the other.
- Transgressive expression: Another subcategory of nonadditive expression, where the *total* homoeologous expression statistically exceeds or falls below both parental levels, termed transgressive up-regulation or transgressive down-regulation, respectively.
- HEB: Unequal expression of homoeologous gene copies of a given gene in allopolyploid or hybrid. The direction of bias is indicated by the homoeolog showing higher expression (e.g. At bias). When the bias direction toward a specific subgenome is preferentially observed *among* genes, the genome-wide HEB becomes "unbalanced."
- *Cis*- and *trans*-regulatory divergence: The regulatory cause of homoeolog gene expression changes relative to parental states (e.g. At/Dt vs. A_2/D_5) can be decomposed into *cis*- and *trans*-acting components. *Cis*-regulatory variants refer to sequence changes in the *cis*-elements near genes, while *trans*-regulatory variants refer to expression or functional changes in *trans*-activating factors like TFs. The relative contributions of *cis* and *trans* variants can be estimated using an ASE framework as illustrated in Fig. 6b.
- Impact of genome evolution: The dynamic changes of relative expression patterns between duplicated genes (e.g. A_2/D_5 in diploids and At/Dt in hybrid and allopolyploid) can be decomposed into effects of hybridization (***Hr***), allopolyploidization (***Pr***), and genome doubling (***Wr***), according to Hu and Wendel (2019) and as formulated in Fig. 6c.

reference genomes, and direct comparison of these profiles revealed a systematic left-ward shift of peak and higher accessibility in A_2 and D_5 around TSSs, regardless of the direction of parental expression divergence (Fig. 7a, top row). Such a pattern is likely due to differences in gene annotation between the two diploid reference genomes, which thereby limits our ability to directly detect accessibility changes associated with parental expression divergence ($\boldsymbol{A} \neq 0$). 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 ($\boldsymbol{Bp} \neq 0$), and higher D- versus A-promoter accessibility was observed for pairs exhibiting D-biases (Fig. 7a, bottom row). Additionally, the homoeolog that exhibited biased higher expression tended to display larger ACRs within 1 kb of the TSS (Fig. 7b). For HEB in F_1 ($\boldsymbol{B} \neq 0$), the homoeologous accessibility profile based on diploid reference genomes (A_2 and D_5 concatenated) exhibited the similar pattern as observed between diploids (Fig. 7a, the second row from top); interestingly, the use of the allopolyploid (AD_1) reference eliminated the positional shift between A- and D-peaks, while the higher A- than D-promoter accessibility level remained, regardless of the direction of HEB (Fig. 7a, middle two rows). Although we cannot rule out artifacts introduced by either reference, the distinct patterns in diploid hybrid versus 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 OGs (referred to as nonOGs: A_2—18850; D_5—14329; At—13227; Dt—15895) and found distinct patterns between OG and nonOG genes (supplementary fig. S9, Supplementary Material online). 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 (Fig. 2) and gene expression (Fig. 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 (A_2, F_1:At, and AD_1:At) and D-genome profiles (D_5, F_1:Dt, and AD_1:Dt) against their corresponding diploid references by genomic content (supplementary figs. S10 and S11, Supplementary Material online). The results revealed a relatively slight decrease by hybridization and a much stronger increase by allopolyploidization in accessibility ($F_1 < A_2/D_5 \ll AD_1$) for both the A- and D-genomes, as evident in aggregation plots (Fig. 7d) and DA tests (Fig. 7c). The DA results also indicated more accessibility increases in At versus Dt promoters, consistent with the previous marginal comparison of ACRs (Fig. 3c and d). In association with the impact of genome doubling on gene expression, the up-regulation of At/Dt homoeolog expression ratios ($\boldsymbol{Wr} > 0$) was attributed to a biased increase in At promoter accessibility, while the down-regulation of At/Dt homoeolog expression ratios ($\boldsymbol{Wr} < 0$) were attributed to a biased increase of Dt promoter accessibility (Fig. 7e). No apparent accessibility patterns were observed with the impact of hybridization ($\boldsymbol{Hr} \neq 0$ in 142 OG; supplementary figs. S10 and S11A, Supplementary Material online), 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 nonadditive patterns accompanying hybridization and polyploidy, we investigated how promoter accessibility changes of At and Dt homoeologs were associated with ELD in F_1 (supplementary fig. S11B, Supplementary Material online) and AD_1 (supplementary fig. S11C, Supplementary Material online); this analysis is summarized in Fig. 7f. In the diploid hybrid, when higher parental expression was detected in A_2 than D_5, 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



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**

Fig. 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, 1,552 genes (6.8% of 22,889 ortholog groups) are more highly expressed in AD_1, and 1,070 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. Nonadditive expression categories include ELD (A-dominant and D-dominant), transgression (up- and down-regulation), and other nonadditive 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 and Wendel 2019), which combined the classic ASE analysis with the estimation of evolutionary impact. d) Summary of parental divergence, HEB, and evolutionary impacts. χ^2 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.

D-dominant ELD. Conversely, when higher parental expression was detected in D_5 compared to A_2, 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 F_1. Interestingly, in allopolyploid cotton, accessibility changes were primarily in At promoters, likely due to sequence evolution accompanying natural allopolyploidization (Fig. 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 S19, Supplementary Material online). 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 A_2, D_5, AD1:At, and AD1:Dt genomes (H1—4, H2A—17, H2B—7, H3—8, and H4—11) (supplementary table S20, Supplementary Material online). We found that the *total* expression of these genes was higher in the allopolyploid (mean sum transcripts per million [TPM] with standard deviation: 4988.4 ± 189.5) compared to the diploid genomes (3714.3 ± 301.1 in F_1, 4338.3 ± 414.9 in D_5, and 3207.2 ± 661.2 in A_2), 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 (Fig. 8a). At the histone variant level (supplementary fig.

S13, Supplementary Material online), this pattern was observed for canonical H1, H2A.X, and H2A.Z variants, due to transgressive up-regulation of their gene members in allopolyploid cotton; interestingly, their counterparts in hybrid F_1 tend to exhibit the D_5-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 (Fig. 8a; supplementary fig. S13, Supplementary Material online). Assuming histone expression levels correlate with the nucleosome number, it is intriguing that the smaller D_5 diploid exhibited statistically equal, or even higher histone expression levels compared to the much larger A_2 diploid (GS 0.8 Gb vs. 1.6 Gb). This finding is consistent with the nucleosome positioning result, i.e. smaller D_5 diploid exhibiting higher NC, suggesting a speculation that histone gene expression may contribute to regulation of nucleosome spacing.

At the OG level (i.e. expression by gene), more nonadditive expressions were detected in the allopolyploid than hybrid (Fig. 8b). Directions of parental divergence (i.e. $\boldsymbol{A} > 0$ and $\boldsymbol{A} < 0$) and HEB (i.e. $\boldsymbol{B} > 0$ and $\boldsymbol{B} < 0$, or $\boldsymbol{Bp} > 0$ and $\boldsymbol{Bp} < 0$) were more or less balanced, which were often influenced by *cis*-only regulation, and no significant *trans*-regulatory divergence was detected (Fig. 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 (Fig. 8d and e: OG0025113—*Gohir.A13G169300* vs. *Gohir.D13G174801*). More examples are shown in supplementary fig. S14, Supplementary Material online.

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: (i) how does genome merger and doubling accompanying allopolyploidy alter chromatin structure; and (ii) 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 ($\boldsymbol{Pr} \gg \boldsymbol{Hr}$), noting that the former entails both genome doubling and, in the case of *Gossypium*, 1 to 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 appears to have been driven by sequence evolution acting in *cis*. First, relative to the parental A_2 and D_5 diploids that model the allopolyploid progenitors, only slight changes of nucleosome organization (Fig. 2b and c) and chromatin accessibility (Fig. 7c) were detected in the F_1 hybrid, with allele-specific patterns closely mirroring those of diploids. This lack of deviation from vertical transmission of preexisting 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 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 (Fig. 2b and c), ACR classification (Table 1), genomic TE content and distribution (Fig. 5a and supplementary fig. S6, Supplementary Material online), accessibility round TSS (Fig. 7a), and the promoter *cis*-regulatory landscape (Fig. 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). Han et al. (2022) also examined DHS distribution patterns in wild versus domesticated *G. hirsutum*, cultivated *G. barbadense*, and wild *G. darwinii*, revealing convergent DHS distributions between subgenomes in all of these allopolyploid cotton genotypes. However, despite the commonalities between At and Dt subgenomes, significant differences in marginal DHS distribution patterns were observed among species. For instance, 12.3% to 13.5% of DHS are genic in *G. hirsutum*, 9.0% to 9.7% in *G. barbadense*, and 31.5% to 32.3% in *G. darwinii*. These interspecific variations that remain to be further characterized underscore the importance of conducting comparative analyses between species and even incorporating multiple accessions from the same species. Recognizing the limitation of our study, which included only one accession each for allopolyploid and diploid genomes, future studies would benefit from including multiple accessions to enhance our understanding of chromatin evolution in cotton.

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 and 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 intrasubgenomic 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 (Fig. 2d and e), where reduced difference in nucleosome spacing by allopolyploidy can be predicted by

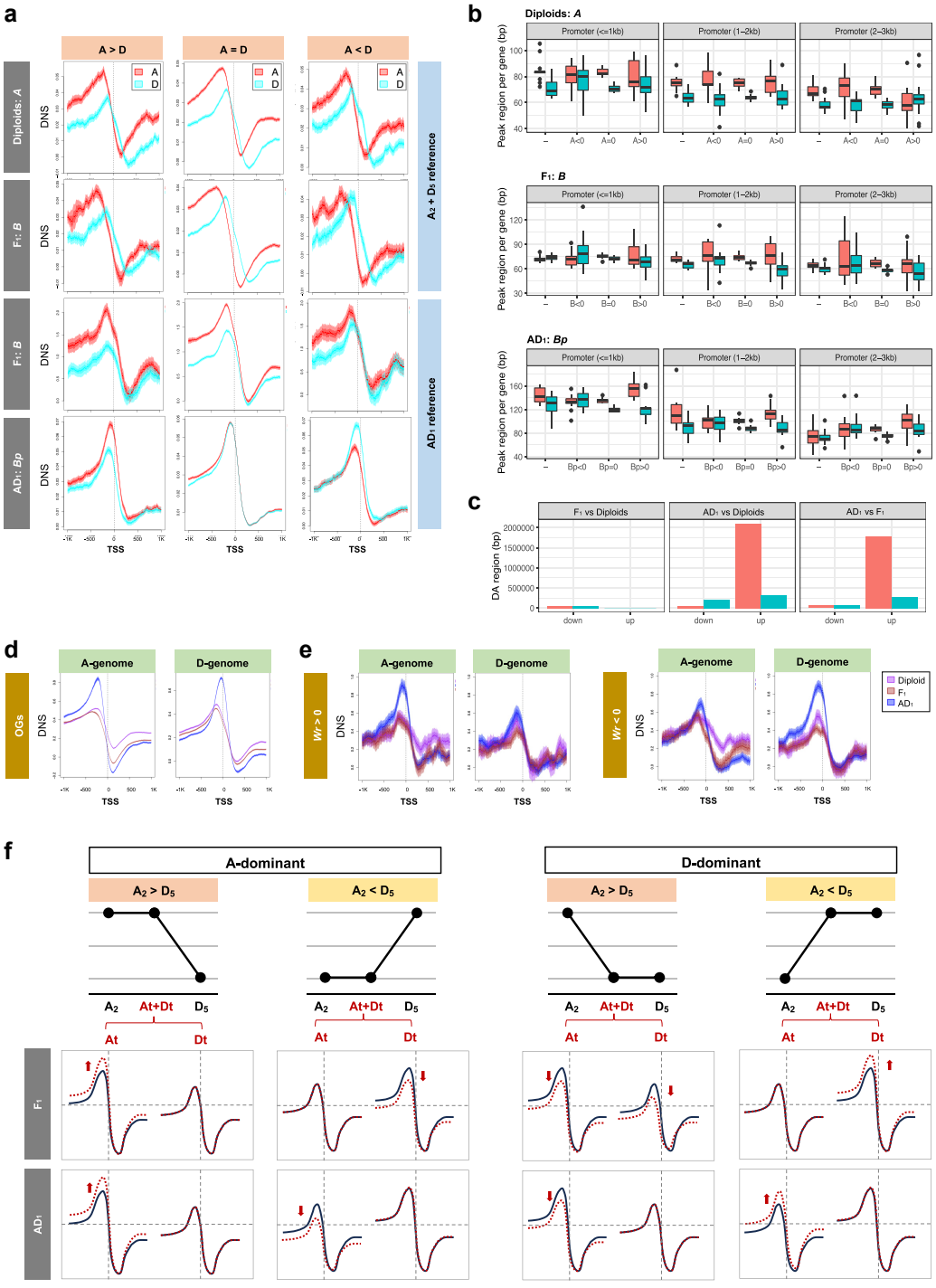


Fig. 7. Promoter accessibility of duplicated genes in diploid and allopolyploid cottons. a) Aggregation plots of DNS signals were present in association with duplicated gene expression patterns of parental divergence (***A*** $\neq 0$; top row), HEBs in F_1 (***B*** $\neq 0$; middle two rows) and in AD_1 (***Bp*** $\neq 0$; bottom row). The *x*-axis is centered on TSS $\pm$ 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. Consistent patterns were also observed from analyzing the chromatin accessibility signals profiled by DNase-seq (Han et al. 2022) as shown in supplementary fig. S12, Supplementary Material online. 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 to 2 kb, and 2 to 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 vs. Dt expression in AD_1, ***Bp*** < 0 indicates higher Dt vs. 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- and D-genomes as color-coded in a). 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.

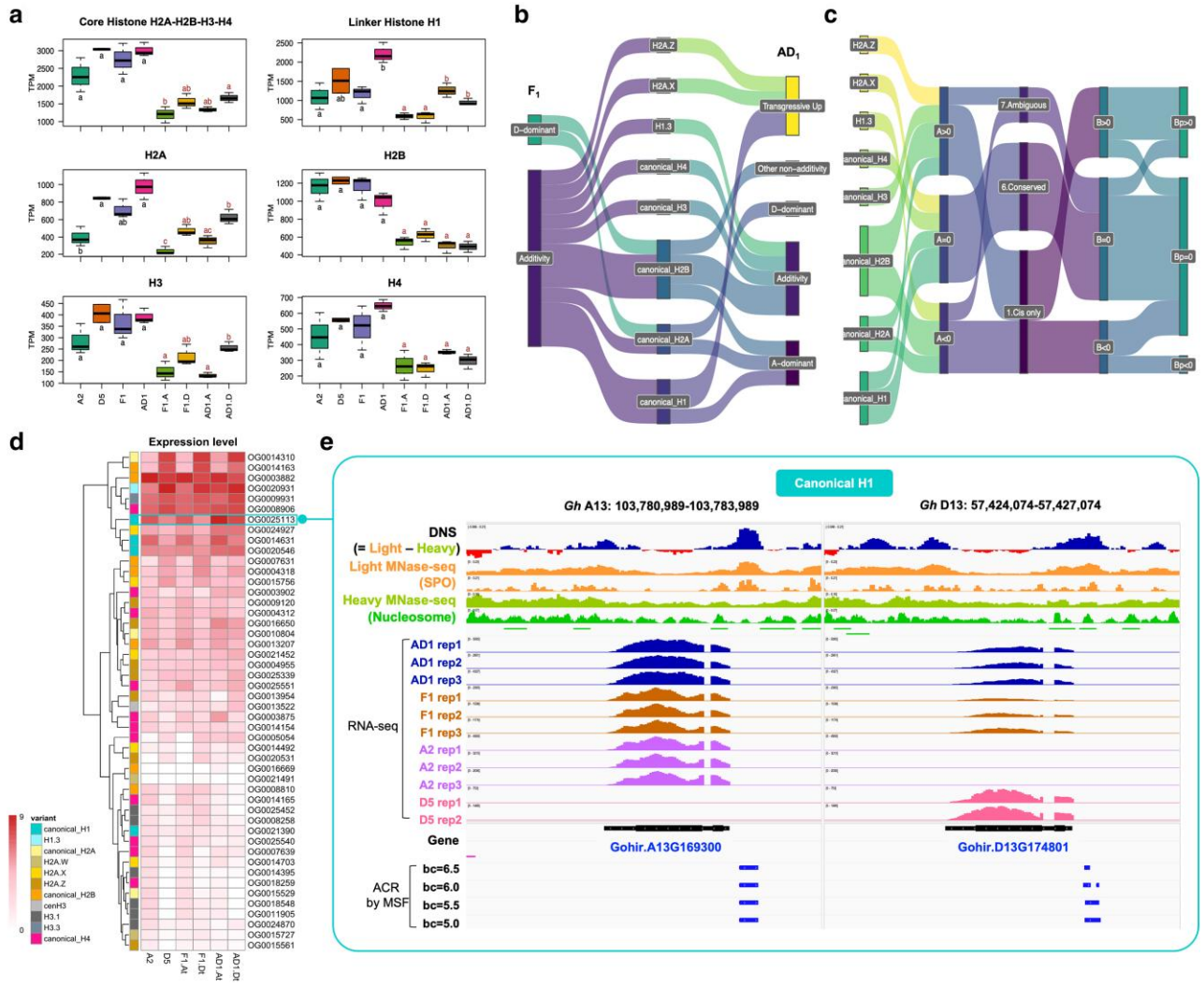


Fig. 8. Analysis of histone gene expression. a) Boxplots present summed expression levels of histone gene family. Comparisons across diploid and allopolyploid cottons. Comparisons between (sub)genomes were performed using ANOVA with post hoc Tukey HSD test ($P < 0.05$). Groups with the same letter are not significantly different. b) The inheritance mode of parental histone expression was compared between F_1 and AD_1, as characterized by additive and nonadditive 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 supplementary fig. S14, Supplementary Material online.

DNA sequence per se. It has been recognized that nucleosome formation favors the 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 and 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 (Fig. 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 noncoding 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 to 185 bp, with shorter NRLs observed in telomeric nucleosomes compared to bulk nucleosomes (Vershinin and 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 were 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) (Zhang et al. 2015; Chen et al. 2017), 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 (Fransz and de Jong 2011; Brouwer et al. 2021): 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, posttranslational 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, 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 and Wendel 2019), which enabled us to first systematically characterize these duplicated gene expression patterns (Fig. 6), and next disentangle the regulatory effects of chromatin accessibility (Fig. 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 (Fig. 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. $\boldsymbol{Wr} > 0$ or $\boldsymbol{Wr} < 0$ in Fig. 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 and Coate 2019) comprise a promising direction of future research.

Nonadditive 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 comprises an interesting research direction. The phenomenon of ELD, another well-known yet mechanistically mysterious nonadditive 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 F_1 generation; in contrast, allopolyploidy is characterized by a distinctive pattern in which accessibility changes predominantly occur in At promoters, a shift likely driven by various biophysical and biochemical factors associated with ploidy stoichiometry, as well as sequence evolution linked to natural allopolyploidization (Fig. 7f). Yoo et al. (2013) previously investigated homoeolog expression levels relative to ELD patterns and also showed that ELD reflects the up- or down-regulation of alternative homoeologs more frequently, compared to the up- or down-regulation 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, Han et al. (2022) examined the relationships between DHS accessibility and the various histone marks, demonstrating the coordinated dynamics among histone modifications, TEs, and DHS landscape under polyploidization. Additionally, 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 F_1, 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 AD_1 (Fig. 6d), the globally higher promoter accessibility in the A- than D-genome remains enigmatic.

The orchestration of 3D 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, and ultimately phenotypic evolution, in cotton and other allopolyploid systems.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Kenneth McCabe for Greenhouse assistance and the Research IT unit at Iowa State University (https://researchit.las.iastate.edu/) for computational support.

Author Contributions

J.F.W., H.W.B., and G.H. conceived the project and designed experiments. E.R.M., J.P.G., and C.E.G. performed RNA-seq experiments. J.A.U. provided ATAC-seq data. G.H. performed MNase-seq experiments, conducted data analyses, with inputs from P.-Y.L., X.S., D.L.V., S.B.G., J.Z., and H.W.B. for MNase-seq, from C.E.G. for RNA-seq, from J.L.C. for ortholog and homoeolog detection, from S.O. for transposable element annotation, from X.X., D.Z., and D.L. for

duplicated gene expression analysis. G.H. wrote the manuscript. C.E.G., H.W.B., and J.F.W. revised the manuscript. All authors read and approved the final manuscript.

Funding

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 J.F.W. and C.E.G., IOS-1444532 to H.W.B. and J.Z., and IOS-1445014 to J.F.W.

Conflict of Interest

The authors declare no competing interests.

References

Adams KL, Cronn R, Percifield R, Wendel JF. Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proc Natl Acad Sci U S A*. 2003:**100**(8):4649–4654. https://doi.org/10.1073/pnas.0630618100.

Ahmad K, Henikoff S, Ramachandran S. Managing the steady state chromatin landscape by nucleosome dynamics. *Ann Rev Biochem*. 2022:**91**(1):183–195. https://doi.org/10.1146/annurev-biochem-032620-104508.

Ando M, Saito Y, Xu G, Bui NQ, Medetgul-Ernar K, Pu M, Fisch K, Ren S, Sakai A, Fukusumi T, *et al.* Chromatin dysregulation and DNA methylation at transcription start sites associated with transcriptional repression in cancers. *Nat Commun*. 2019:**10**(1):2188. https://doi.org/10.1038/s41467-019-09937-w.

Andrews AJ, Luger K. Nucleosome structure(s) and stability: variations on a theme. *Ann Rev Biophys*. 2011:**40**(1):99–117. https://doi.org/10.1146/annurev-biophys-042910-155329.

Arrigo N, de La Harpe M, Litsios G, Zozomová-Lihová J, Španiel S, Marhold K, Barker MS, Alvarez N. Is hybridization driving the evolution of climatic niche in *Alyssum montanum*. *Am J Bot*. 2016:**103**(7):1348–1357. https://doi.org/10.3732/ajb.1500368.

Arya G, Maitra A, Grigoryev SA. A structural perspective on the where, how, why, and what of nucleosome positioning. *J Biomol Struct Dyn*. 2010:**27**(6):803–820. https://doi.org/10.1080/07391102.2010.10508585.

Athey BD, Smith MF, Rankert DA, Williams SP, Langmore JP. The diameters of frozen-hydrated chromatin fibers increase with DNA linker length: evidence in support of variable diameter models for chromatin. *J Cell Biol*. 1990:**111**(3):795–806. https://doi.org/10.1083/jcb.111.3.795.

Bailey TL, Grant CE. SEA: simple enrichment analysis of motifs. bioRxiv 457422. https://doi.org/10.1101/2021.08.23.457422., 24 August 2021, preprint: not peer reviewed.

Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. *Nucleic Acids Res*. 2015:**43**(W1):W39–W49. https://doi.org/10.1093/nar/gkv416.

Baldi S, Korber P, Becker PB. Beads on a string—nucleosome array arrangements and folding of the chromatin fiber. *Nature Struct Mol Biol*. 2020:**27**(2):109–118. https://doi.org/10.1038/s41594-019-0368-x.

Baniaga AE, Marx HE, Arrigo N, Barker MS. Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecol Lett*. 2020:**23**(1):68–78. https://doi.org/10.1111/ele.13402.

Bao Y, Hu G, Grover CE, Conover J, Yuan D, Wendel JF. Unraveling cis and trans regulatory evolution during cotton domestication. *Nat Commun*. 2019:**10**(1):5399. https://doi.org/10.1038/s41467-019-13386-w.

Barbier J, Vaillant C, Volff J-N, Brunet FG, Audit B. Coupling between sequence-mediated nucleosome organization and genome evolution. *Genes (Basel)*. 2021:**12**(6):851. https://doi.org/10.3390/genes12060851.

Bottani S, Zabet NR, Wendel JF, Veitia RA. Gene expression dominance in allopolyploids: hypotheses and models. *Trends Plant Sci*. 2018:**23**(5):393–402. https://doi.org/10.1016/j.tplants.2018.01.002.

Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-Seq quantification. *Nat Biotechnol*. 2016:**34**(5):525–527. https://doi.org/10.1038/nbt.3519.

Brouwer T, Pham C, Kaczmarczyk A, de Voogd W-J, Botto M, Vizjak P, Mueller-Planitz F, van Noort J. A critical role for linker DNA in higher-order folding of chromatin fibers. *Nucleic Acids Res*. 2021:**49**(5):2537–2551. https://doi.org/10.1093/nar/gkab058.

Bubb KL, Deal RB. Considerations in the analysis of plant chromatin accessibility data. *Curr Opin Plant Biol*. 2020:**54**:69–78. https://doi.org/10.1016/j.pbi.2020.01.003.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*. 2013:**10**(12):1213–1218. https://doi.org/10.1038/nmeth.2688.

Buggs RJ, Wendel JF, Doyle JJ, Soltis DE, Soltis PS, Coate JE. The legacy of diploid progenitors in allopolyploid gene expression patterns. *Philos Trans R Soc Lond B Biol Sci*. 2014:**369**(1648):20130354. https://doi.org/10.1098/rstb.2013.0354.

Castro-Mondragon JA, Jaeger S, Thieffry D, Thomas-Chollier M, van Helden J. RSAT matrix-clustering: dynamic exploration and redundancy reduction of transcription factor binding motif collections. *Nucleic Acids Res*. 2017:**45**(13):e119. https://doi.org/10.1093/nar/gkx314.

Chen Z, Grover CE, Li P, Wang Y, Nie H, Zhao Y, Wang M, Liu F, Zhou Z, Wang X, *et al.* Molecular evolution of the plastid genome during diversification of the cotton genus. *Mol Phylogenet Evol*. 2017:**112**:268–276. https://doi.org/10.1016/j.ympev.2017.04.014.

Chen J, Li E, Zhang X, Dong X, Lei L, Song W, Zhao H, Lai J. Genome-wide nucleosome occupancy and organization modulates the plasticity of gene transcriptional status in maize. *Mol Plant*. 2017:**10**(7):962–974. https://doi.org/10.1016/j.molp.2017.05.001.

Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, *et al.* Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet*. 2020:**52**(5):525–533. https://doi.org/10.1038/s41588-020-0614-5.

Chodavarapu RK, Feng S, Bernatavichute YV, Chen P-Y, Stroud H, Yu Y, Hetzel JA, Kuo F, Kim J, Cokus SJ, *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature*. 2010:**466**(7304):388–392. https://doi.org/10.1038/nature09147.

Coate JE, Luciano AK, Seralathan V, Minchew KJ, Owens TG, Doyle JJ. Anatomical, biochemical, and photosynthetic responses to recent allopolyploidy in *Glycine dolichocarpa* (Fabaceae). *Am J Bot*. 2012:**99**(1):55–67. https://doi.org/10.3732/ajb.1100465.

Combes M-C, Joët T, Stavrinides AK, Lashermes P. New cup out of old coffee: contribution of parental gene expression legacy to phenotypic novelty in coffee beans of the allopolyploid *Coffea arabica* L. *Ann Bot*. 2022:**131**(1):157–170. https://doi.org/10.1093/aob/mcac041.

Conover JL, Sharbrough J, Wendel JF. pSONIC: ploidy-aware syntenic orthologous networks identified via collinearity. *G3 (Bethesda)*. 2021:**11**(8):jkab170. https://doi.org/10.1093/g3journal/jkab170.

Conover JL, Wendel JF. Deleterious mutations accumulate faster in allopolyploid than diploid cotton (*Gossypium*) and unequally between subgenomes. *Mol Biol Evol*. 2022:**39**(2):msac024. https://doi.org/10.1093/molbev/msac024.

Coughlan JM, Han S, Stefanović S, Dickinson TA. Widespread generalist clones are associated with range and niche expansion in allopolyploids of Pacific Northwest Hawthorns (*Crataegus* L). *Mol Ecol*. 2017:**26**(20):5484–5499. https://doi.org/10.1111/mec.14331.

Dorrity MW, Alexandre CM, Hamm MO, Vigil A-L, Fields S, Queitsch C, Cuperus JT. The regulatory landscape of *Arabidopsis thaliana* roots at single-cell resolution. *Nat Commun*. 2021:**12**(1):3334. https://doi.org/10.1038/s41467-021-23675-y.

Doyle JJ, Coate JE. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *Int J Plant Sci*. 2019:**180**(1):1–52. https://doi.org/10.1086/700636.

Draizen EJ, Shaytan AK, Mariño-Ramírez L, Talbert PB, Landsman D, Panchenko AR. HistoneDB 2.0: a histone database with variants—an integrated resource to explore histones and their variants. *Database (Oxford)*. 2016:**2016**:baw014. https://doi.org/10.1093/database/baw014.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004:**32**(5):1792–1797. https://doi.org/10.1093/nar/gkh340.

Elliott TL, Muasya M, Bureš P. Complex patterns of ploidy in a holocentric plant clade (Schoenus, Cyperaceae) in the Cape biodiversity hotspot. *Ann Bot*. 2023:**131**(1):143–156. https://doi.org/10.1093/aob/mcac027.

Fang C, Yang M, Tang Y, Zhang L, Zhao H, Ni H, Chen Q, Meng F, Jiang J. Dynamics of *cis*-regulatory sequences and transcriptional divergence of duplicated genes in soybean. *Proc Natl Acad Sci U S A*. 2023:**120**(44):e2303836120. https://doi.org/10.1073/pnas.2303836120.

Flagel L, Udall J, Nettleton D, Wendel J. Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution. *BMC Biol*. 2008:**6**(1):16. https://doi.org/10.1186/1741-7007-6-16.

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

Fox DT, Soltis DE, Soltis PS, Ashman T-L, Van de Peer Y. Polyploidy: a biological force from cells to ecosystems. *Trends Cell Biol*. 2020:**30**(9):688–694. https://doi.org/10.1016/j.tcb.2020.06.006.

Fransz P, de Jong H. From nucleosome to chromosome: a dynamic organization of genetic information. *Plant J*. 2011:**66**(1):4–17. https://doi.org/10.1111/j.1365-313X.2011.04526.x.

Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. Controls of nucleosome positioning in the human genome. *PLoS Genet*. 2012:**8**(11):e1003036. https://doi.org/10.1371/journal.pgen.1003036.

Gallagher JP, Grover CE, Hu G, Jareczek JJ, Wendel JF. Conservation and divergence in duplicated fiber coexpression networks accompanying domestication of the polyploid *Gossypium hirsutum* L. *G3 (Bethesda)*. 2020:**10**(8):2879–2892. https://doi.org/10.1534/g3.120.401362.

Gallagher JP, Grover CE, Hu G, Wendel JF. Insights into the ecology and evolution of polyploid plants through network analysis. *Mol Ecol*. 2016:**25**(11):2644–2660. https://doi.org/10.1111/mec.13626.

Gallagher JP, Grover CE, Rex K, Moran M, Wendel JF. A new species of cotton from Wake Atoll, *Gossypium stephensii* (Malvaceae). *Systematic Bot*. 2017:**42**(1):115–123. https://doi.org/10.1600/036364417X694593.

Galli M, Feng F, Gallavotti A. Mapping regulatory determinants in plants. *Front Genet*. 2020:**11**:591194. https://doi.org/10.3389/fgene.2020.591194.

Giles KA, Taberlay PC. The role of nucleosomes in epigenetic gene regulation. In Hesson LB, Pritchard AL, editors. *Clinical epigenetics*. Singapore: Springer; 2019. pp. 87–117.

Giraud D, Lima O, Rousseau-Gueutin M, Salmon A, Aïnouche M. Gene and transposable element expression evolution following recent and past polyploidy events in *Spartina* (Poaceae). *Front Genet*. 2021:**12**:589160. https://doi.org/10.3389/fgene.2021.589160.

Girimurugan SB, Liu Y, Lung P-Y, Vera DL, Dennis JH, Bass HW, Zhang J. Iseg: an efficient algorithm for segmentation of genomic and epigenomic data. *BMC Bioinformatics*. 2018:**19**(1):131. https://doi.org/10.1186/s12859-018-2140-3.

Gouy M, Tannier E, Comte N, Parsons DP. Seaview version 5: a multiplatform software for multiple sequence alignment, molecular phylogenetic analyses, and tree reconciliation. *Methods Mol Biol*. 2021:**2231**:241–260. https://doi.org/10.1007/978-1-0716-1036-7_15.

Grant CE, Bailey TL. XSTREME: Comprehensive motif analysis of biological sequence datasets. bioRxiv 458722, https://doi.org/10.1101/2021.09.02.458722. September 03, 2021, preprint: not peer reviewed.

Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011:**27**(7):1017–1018. https://doi.org/10.1093/bioinformatics/btr064.

Grossman SR, Engreitz J, Ray JP, Nguyen TH, Hacohen N, Lander ES. Positional specificity of different transcription factor classes within enhancers. *Proc Natl Acad Sci U S A*. 2018:**115**(30):E7222–E7230. https://doi.org/10.1073/pnas.1804663115.

Grover CE, Gallagher JP, Szadkowski EP, Yoo MJ, Flagel LE, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol*. 2012:**196**(4):966–971. https://doi.org/10.1111/j.1469-8137.2012.04365.x.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010:**59**(3):307–321. https://doi.org/10.1093/sysbio/syq010.

Han J, Lopez-Arredondo D, Yu G, Wang Y, Wang B, Wall SB, Zhang X, Fang H, Barragán-Rosillo AC, Pan X, *et al.* Genome-wide chromatin accessibility analysis unveils open chromatin convergent evolution during polyploidization in cotton. *Proc Natl Acad Sci U S A*. 2022:**119**(44):e2209743119. https://doi.org/10.1073/pnas.2209743119.

Han J, Wang P, Wang Q, Lin Q, Chen Z, Yu G, Miao C, Dao Y, Wu R, Schnable JC, *et al.* Genome-wide characterization of DNase I-hypersensitive sites and cold response regulatory landscapes in grasses. *Plant Cell*. 2020:**32**(8):2457–2473. https://doi.org/10.1105/tpc.19.00716.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010:**38**(4):576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Heslop-Harrison JSP, Schwarzacher T, Liu Q. Polyploidy: its consequences and enabling role in plant diversification and evolution. *Ann Bot*. 2022:**131**(1):1–10. https://doi.org/10.1093/aob/mcac132.

Hofmeister BT, Lee K, Rohr NA, Hall DW, Schmitz RJ. Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation. *Genome Biol*. 2017:**18**(1):155. https://doi.org/10.1186/s13059-017-1288-x.

Hu G, Grover CE, Yuan D, Dong Y, Miller E, Conover JL, Wendel JF. Evolution and diversity of the cotton genome. In: Rahman MU, Zafar Y, Zhang T, editors. *Cotton precision breeding*. Springer International Publishing; 2021. p. 25–78.

Hu G, Hovav R, Grover CE, Faigenboim-Doron A, Kadmon N, Page JT, Udall JA, Wendel JF. Evolutionary conservation and divergence of gene coexpression networks in Gossypium (cotton) seeds. *Genome Biol Evol*. 2016:**8**(12):3765–3783. https://doi.org/10.1093/gbe/evw280.

Hu G, Koh J, Yoo M-J, Chen S, Wendel JF. Gene-expression novelty in allopolyploid cotton: a proteomic perspective. *Genetics*. 2015:**200**(1):91–104. https://doi.org/10.1534/genetics.115.174367.

Hu G, Koh J, Yoo M-J, Grupp K, Chen S, Wendel JF. Proteomic profiling of developing cotton fibers from wild and domesticated *Gossypium barbadense*. *New Phytol*. 2013:**200**(2):570–582. https://doi.org/10.1111/nph.12381.

Hu G, Koh J, Yoo M-J, Pathak D, Chen S, Wendel JF. Proteomics profiling of fiber development and domestication in upland cotton

(*Gossypium hirsutum* L.). *Planta* 2014:**240**(6):1237–1251. https://doi.org/10.1007/s00425-014-2146-7.

Hu G, Wendel JF. Cis-trans controls and regulatory novelty accompanying allopolyploidization. *New Phytol*. 2019:**221**(4):1691–1700. https://doi.org/10.1111/nph.15515.

Huang G, Huang J-Q, Chen X-Y, Zhu Y-X. Recent advances and future perspectives in cotton research. *Annu Rev Plant Biol*. 2021:**72**(1):437–462. https://doi.org/10.1146/annurev-arplant-080720-113241.

Huang G, Wu Z, Percy RG, Bai M, Li Y, Frelichowski JE, Hu J, Wang K, Yu JZ, Zhu Y. Genome sequence of *Gossypium herbaceum* and genome updates of *Gossypium arboreum* and *Gossypium hirsutum* provide insights into cotton A-genome evolution. *Nat Genet*. 2020:**52**(5):516–524. https://doi.org/10.1038/s41588-020-0607-4.

Initiative OTPT, One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 2019:**574**(7780): 679–685. https://doi.org/10.1038/s41586-019-1693-2.

Jackson SA. Epigenomics: dissecting hybridization and polyploidization. *Genome Biol*. 2017:**18**(1):117. https://doi.org/10.1186/s13059-017-1254-7.

Jiang J. The "dark matter" in the plant genomes: non-coding and unannotated DNA sequences associated with open chromatin. *Curr Opin Plant Biol*. 2015:**24**:17–23. https://doi.org/10.1016/j.pbi.2015.01.005.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, *et al.* Ancestral polyploidy in seed plants and angiosperms. *Nature*. 2011:**473**(7345):97–100. https://doi.org/10.1038/nature09916.

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res*. 2017:**45**(D1):D1040–D1045. https://doi.org/10.1093/nar/gkw982.

Jordan KW, He F, de Soto MF, Akhunova A, Akhunov E. Differential chromatin accessibility landscape reveals structural and functional features of the allopolyploid wheat chromosomes. *Genome Biol*. 2020:**21**(1):176. https://doi.org/10.1186/s13059-020-02093-1.

Jordan WT, Schmitz RJ. The shocking consequences of hybrid epigenomes [review of *The shocking consequences of hybrid epigenomes*]. *Genome Biol*. 2016:**17**(1):85. https://doi.org/10.1186/s13059-016-0967-3.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*. 2009:**458**(7236):362–366. https://doi.org/10.1038/nature07667.

Kawakatsu T, Huang S-SC, Jupe F, Sasaki E, Schmitz RJ, Urich MA, Castanon R, Nery JR, Barragan C, He Y, *et al.* Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell*. 2016:**166**(2):492–505. https://doi.org/10.1016/j.cell.2016.06.044.

Klein DC, Hainer SJ. Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Res*. 2020:**28**(1):69–85. https://doi.org/10.1007/s10577-019-09619-9.

Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet*. 2019:**20**(4):207–220. https://doi.org/10.1038/s41576-018-0089-8.

Knight CA, Beaulieu JM. Genome size scaling through phenotype space. *Ann Bot*. 2008:**101**(6):759–766. https://doi.org/10.1093/aob/mcm321.

Kolde R. Pheatmap: pretty heatmaps. *R Package Version*. 2019:**1**(2):726.

Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974:**184**(4139):868–871. https://doi.org/10.1126/science.184.4139.868.

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. https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. [Accessed 2016 April 28].

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 2012:**22**(9):1813–1831. https://doi.org/10.1101/gr.136184.111.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012:**9**(4):357–359. https://doi.org/10.1038/nmeth.1923.

Lantermann AB, Straub T, Strålfors A, Yuan G-C, Ekwall K, Korber P. *Schizosaccharomyces pombe* genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of *Saccharomyces cerevisiae*. *Nat Struct Mol Biol*. 2010:**17**(2):251–257. https://doi.org/10.1038/nsmb.1741.

Leitch AR, Leitch IJ. Genomic plasticity and the diversity of polyploid plants. *Science*. 2008:**320**(5875):481–483. https://doi.org/10.1126/science.1153585.

Levin DA. Polyploidy and novelty in flowering plants. *Am Nat*. 1983:**122**(1):1–25. https://doi.org/10.1086/284115.

Li G, Liu S, Wang J, He J, Huang H, Zhang Y, Xu L. ISWI proteins participate in the genome-wide nucleosome distribution in *Arabidopsis*. *Plant J*. 2014:**78**(4):706–714. https://doi.org/10.1111/tpj.12499.

Liang Z, Myers ZA, Petrella D, Engelhorn J, Hartwig T, Springer NM. Mapping responsive genomic elements to heat stress in a maize diversity panel. *Genome Biol*. 2022:**23**(1):234. https://doi.org/10.1186/s13059-022-02807-7.

Liu M-J, Seddon AE, Tsai ZT-Y, Major IT, Floer M, Howe GA, Shiu S-H. Determinants of nucleosome positioning and their influence on plant gene expression. *Genome Res*. 2015:**25**(8):1182–1195. https://doi.org/10.1101/gr.188680.114.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014:**15**(12):550. https://doi.org/10.1186/s13059-014-0550-8.

Lu Z, Hofmeister BT, Vollmers C, DuBois RM, Schmitz RJ. Combining ATAC-seq with nuclei sorting for discovery of *cis*-regulatory regions in plant genomes. *Nucleic Acids Res*. 2017:**45**(6):e41. https://doi.org/10.1093/nar/gkw1179.

Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ. The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat Plants*. 2019:**5**(12):1250–1259. https://doi.org/10.1038/s41477-019-0548-z.

Luger K, Rechsteiner TJ, Flaus AJ, Waye MM, Richmond TJ. Characterization of nucleosome core particles containing histone proteins made in bacteria. *J Mol Biol*. 1997:**272**(3):301–311. https://doi.org/10.1006/jmbi.1997.1235.

Lun ATL, Smyth GK. Csaw: a Bioconductor package for differential binding analysis of ChIP-seq data using sliding windows. *Nucleic Acids Res*. 2016:**44**(5):e45. https://doi.org/10.1093/nar/gkv1191.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMB Net J*. 2011:**17**(1):10. https://doi.org/10.14806/ej.17.1.200.

Mata JK, Martin SL, Smith TW. Global biodiversity data suggest allopolyploid plants do not occupy larger ranges or harsher conditions compared with their progenitors. *Ecol Evol*. 2023:**13**(8):e10231. https://doi.org/10.1002/ece3.10231.

McLeay RC, Bailey TL. Motif enrichment analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics*. 2010:**11**(1):165. https://doi.org/10.1186/1471-2105-11-165.

Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ, *et al.* MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun*. 2016:**7**(1):11485. https://doi.org/10.1038/ncomms11485.

Mishra MK. Stomatal characteristics at different ploidy levels in *Coffea* L. *Ann Bot*. 1997:**80**(5):689–692. https://doi.org/10.1006/anbo.1997.0491.

Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA, *et al.* Widespread natural

variation of DNA methylation within angiosperms. *Genome Biol*. 2016:**17**(1):194. https://doi.org/10.1186/s13059-016-1059-0.
Olsen KM, Wendel JF. A bountiful harvest: genomic insights into crop domestication phenotypes. *Ann Rev Plant Biol*. 2013:**64**(1):47–70. https://doi.org/10.1146/annurev-arplant-050312-120048.
Orr-Weaver TL. When bigger is better: the role of polyploidy in organogenesis. *Trends Genet*. 2015:**31**(6):307–315. https://doi.org/10.1016/j.tig.2015.03.011.
Otto SP. The evolutionary consequences of polyploidy. *Cell*. 2007:**131**(3):452–462. https://doi.org/10.1016/j.cell.2007.10.022.
Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol*. 2019:**20**(1):275. https://doi.org/10.1186/s13059-019-1905-y.
Page JT, Gingle AR, Udall JA. PolyCat: a resource for genome categorization of sequencing reads from allopolyploid organisms. *G3 (Bethesda)*. 2013:**3**(3):517–525. https://doi.org/10.1534/g3.112.005298.
Parshuram ZA, Harrison TL, Simonsen AK, Stinchcombe JR, Frederickson ME. Nonsymbiotic legumes are more invasive, but only if polyploid. *New Phytol*. 2022:**237**(3):758–765. https://doi.org/10.1111/nph.18579.
Parvathaneni RK, Bertolini E, Shamimuzzaman M, Vera DL, Lung P-Y, Rice BR, Zhang J, Brown PJ, Lipka AE, Bass HW, *et al.* The regulatory landscape of early maize inflorescence development. *Genome Biol*. 2020:**21**(1):165. https://doi.org/10.1186/s13059-020-02070-8.
Pass DA, Sornay E, Marchbank A, Crawford MR, Paszkiewicz K, Kent NA, Murray JAH. Genome-wide chromatin mapping with size resolution reveals a dynamic sub-nucleosomal landscape in *Arabidopsis*. *PLoS Genet*. 2017:**13**(9):e1006988. https://doi.org/10.1371/journal.pgen.1006988.
Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature*. 2012:**492**(7429):423–427. https://doi.org/10.1038/nature11798.
Pei L, Li G, Lindsey K, Zhang X, Wang M. Plant 3D genomics: the exploration and application of chromatin organization. *New Phytol*. 2021:**230**(5):1772–1786. https://doi.org/10.1111/nph.17262.
Probst AV, Desvoyes B, Gutierrez C. Similar yet critically different: the distribution, dynamics and function of histone variants. *J Exp Bot*. 2020:**71**(17):5191–5204. https://doi.org/10.1093/jxb/eraa230.
Qiu Z, Li R, Zhang S, Wang K, Xu M, Li J, Du Y, Yu H, Cui X. Identification of regulatory DNA elements using genome-wide mapping of DNase I hypersensitive sites during tomato fruit development. *Mol Plant*. 2016:**9**(8):1168–1182. https://doi.org/10.1016/j.molp.2016.05.013.
Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. *Curr Protoc Bioinformatics*. 2014:**47**(1):11.12.1–11.12.34. https://doi.org/10.1002/0471250953.bi1112s47.
Radman-Livaja M, Rando OJ. Nucleosome positioning: how is it established, and why does it matter? *Dev Biol*. 2010:**339**(2):258–266. https://doi.org/10.1016/j.ydbio.2009.06.012.
Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014:**42**(W1):W187–W191. https://doi.org/10.1093/nar/gku365.
Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016:**44**(W1):W160–W165. https://doi.org/10.1093/nar/gkw257.
Ramsey J, Schemske DW. Neopolyploidy in flowering plants. *Ann Rev Ecol Sys*. 2002:**33**(1):589–639. https://doi.org/10.1146/annurev.ecolsys.33.010802.150437.
Rapp RA, Udall JA, Wendel JF. Genomic expression dominance in allopolyploids. *BMC Biol*. 2009:**7**(1):18. https://doi.org/10.1186/1741-7007-7-18.
Renny-Byfield S, Wendel JF. Doubling down on genomes: polyploidy and crop plants. *Am J Bot*. 2014:**101**(10):1711–1725. https://doi.org/10.3732/ajb.1400119.
Reske JJ, Wilson MR, Chandler RL. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin*. 2020:**13**(1):22. https://doi.org/10.1186/s13072-020-00342-y.
Reynoso MA, Borowsky AT, Pauluzzi GC, Yeung E, Zhang J, Formentin E, Velasco J, Cabanlit S, Duvenjian C, Prior MJ, *et al.* Gene regulatory networks shape developmental plasticity of root cell types under water extremes in rice. *Dev Cell*. 2022:**57**(9):1177–1192.e6. https://doi.org/10.1016/j.devcel.2022.04.013.
Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, *et al.* Widespread long-range *cis*-regulatory elements in the maize genome. *Nat Plants*. 2019:**5**(12):1237–1249. https://doi.org/10.1038/s41477-019-0547-0.
Robinson MD, McCarthy DJ, Smyth GK. Edger: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010:**26**(1):139–140. https://doi.org/10.1093/bioinformatics/btp616.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011:**29**(1):24–26. https://doi.org/10.1038/nbt.1754.
Rodgers-Melnick E, Vera DL, Bass HW, Buckler ES. Open chromatin reveals the functional maize genome. *Proc Natl Acad Sci U S A*. 2016:**113**(22):E3177–E3184. https://doi.org/10.1073/pnas.1525244113.
Ruprecht C, Lohaus R, Vanneste K, Mutwil M, Nikoloski Z, Van de Peer Y, Persson S. Revisiting ancestral polyploidy in plants. *Sci Adv*. 2017:**3**(7):e1603195. https://doi.org/10.1126/sciadv.1603195.
Savadel SD, Hartwig T, Turpin ZM, Vera DL, Lung P-Y, Sui X, Blank M, Frommer WB, Dennis JH, Zhang J, *et al.* The native cistrome and sequence motif families of the maize ear. *PLoS Genet*. 2021:**17**(8):e1009689. https://doi.org/10.1371/journal.pgen.1009689.
Schmitz RJ, Marand AP, Zhang X, Mosher RA, Turck F, Chen X, Axtell MJ, Zhong X, Brady SM, Megraw M, *et al.* Quality control and evaluation of plant epigenomics data. *Plant Cell*. 2022:**34**(1):503–513. https://doi.org/10.1093/plcell/koab255.
Schmitz RJ, Schultz MD, Lewsey MG, O'Malley RC, Urich MA, Libiger O, Schork NJ, Ecker JR. Transgenerational epigenetic instability is a source of novel methylation variants. *Science*. 2011:**334**(6054):369–373. https://doi.org/10.1126/science.1212959.
Schnable JC, Springer NM, Freeling M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci U S A*. 2011:**108**(10):4069–4074. https://doi.org/10.1073/pnas.1101368108.
Segal E, Widom J. What controls nucleosome positions? *Trends Genets*. 2009:**25**(8):335–343. https://doi.org/10.1016/j.tig.2009.06.002.
Shan S, Boatwright JL, Liu X, Chanderbali AS, Fu C, Soltis PS, Soltis DE. Transcriptome dynamics of the inflorescence in reciprocally formed allopolyploid *Tragopogon miscellus* (Asteraceae). *Front Genets*. 2020:**11**:888. https://doi.org/10.3389/fgene.2020.00888.
Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol*. 2016:**30**:159–165. https://doi.org/10.1016/j.pbi.2016.03.015.
Soman A, Wong SY, Korolev N, Surya W, Lattmann S, Vogirala VK, Chen Q, Berezhnoy NV, van Noort J, Rhodes D, *et al.* Columnar structure of human telomeric chromatin. *Nature*. 2022:**609**(7929):1048–1055. https://doi.org/10.1038/s41586-022-05236-5.
Song Q, Zhang T, Stelly DM, Chen ZJ. Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol*. 2017:**18**(1):99. https://doi.org/10.1186/s13059-017-1229-8.
Springer NM, Schmitz RJ. Exploiting induced and natural epigenetic variation for crop improvement. *Nat Rev Genet*. 2017:**18**(9):563–575. https://doi.org/10.1038/nrg.2017.45.
Stebbins GL. The significance of polyploidy in plant evolution. *Am Nat*. 1940:**74**(750):54–66. https://doi.org/10.1086/280872.

Stewart JM, Craven LA, Brubaker C, Wendel JF. *Gossypium anapoides* (Malvaceae), a new species from western *Australia*. *Novon: J Bot Nomencl*. 2015:**23**(4):447–451. https://doi.org/10.3417/2007140.

Sugiyama S-I. Polyploidy and cellular mechanisms changing leaf size: comparison of diploid and autotetraploid populations in two species of Lolium. *Ann Bot*. 2005:**96**(5):931–938. https://doi.org/10.1093/aob/mci245.

Sullivan AM, Bubb KL, Sandstrom R, Stamatoyannopoulos JA, Queitsch C. DNase I hypersensitivity mapping, genomic footprinting, and transcription factor networks in plants. *Curr Plant Biol*. 2015:**3-4**:40–47. https://doi.org/10.1016/j.cpb.2015.10.001.

Talbert PB, Meers MP, Henikoff S. Old cogs, new tricks: the evolution of gene expression in a chromatin context. *Nat Rev Genet*. 2019:**20**(5):283–297. https://doi.org/10.1038/s41576-019-0105-7.

Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021:**38**(7):3022–3027. https://doi.org/10.1093/molbev/msab120.

Teves SS, Henikoff S. Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. *Genes Dev*. 2011:**25**(22):2387–2397. https://doi.org/10.1101/gad.177675.111.

Tommerup H, Dousmanis A, de Lange T. Unusual chromatin in human telomeres. *Mol Cell Biol*. 1994:**14**(9):5777–5785. https://doi.org/10.1128/mcb.14.9.5777-5785.1994.

Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*. 2014:**7**(1):33. https://doi.org/10.1186/1756-8935-7-33.

Turpin ZM, Vera DL, Savadel SD, Lung P-Y, Wear EE, Mickelson-Young L, Thompson WF, Hanley-Bowdoin L, Dennis JH, Zhang J, *et al*. Chromatin structure profile data from DNS-seq: differential nuclease sensitivity mapping of four reference tissues of B73 maize (*Zea mays* L). *Data Brief*. 2018:**20**:358–363. https://doi.org/10.1016/j.dib.2018.08.015.

Vainshtein Y, Rippe K, Teif VB. NucTools: analysis of chromatin feature occupancy profiles from high-throughput sequencing data. *BMC Genomics*. 2017:**18**(1):158. https://doi.org/10.1186/s12864-017-3580-2.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. Determinants of nucleosome organization in primary human cells. *Nature*. 2011:**474**(7352):516–520. https://doi.org/10.1038/nature10002.

Van de Peer Y, Ashman T-L, Soltis PS, Soltis DE. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*. 2021:**33**(1):11–26. https://doi.org/10.1093/plcell/koaa015.

Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet*. 2017:**18**(7):411–424. https://doi.org/10.1038/nrg.2017.26.

Vera DL, Madzima TF, Labonne JD, Alam MP, Hoffman GG, Girimurugan SB, Zhang J, McGinnis KM, Dennis JH, Bass HW. Differential nuclease sensitivity profiling of chromatin reveals biochemical footprints coupled to gene expression and functional DNA elements in maize. *Plant Cell*. 2014:**26**(10):3883–3893. https://doi.org/10.1105/tpc.114.130609.

Vershinin AV, Heslop-Harrison JS. Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol Biol*. 1998:**36**(1):149–161. https://doi.org/10.1023/A:1005912822671.

Viot CR, Wendel JF. Evolution of the cotton genus, *Gossypium*, and its domestication in the Americas. *Criti Rev Plant Sci*. 2023:**42**(1):1–33. https://doi.org/10.1080/07352689.2022.2156061.

Visger CJ, Wong GK-S, Zhang Y, Soltis PS, Soltis DE. Divergent gene expression levels between diploid and autotetraploid Tolmiea relative to the total transcriptome, the cell, and biomass. *Am J Bot*. 2019:**106**(2):280–291. https://doi.org/10.1002/ajb2.1239.

Voong LN, Xi L, Wang J-P, Wang X. Genome-wide mapping of the nucleosome landscape by micrococcal nuclease and chemical mapping. *Trends Genet*. 2017:**33**(8):495–507. https://doi.org/10.1016/j.tig.2017.05.007.

Wang M, Tu L, Lin M, Lin Z, Wang P, Yang Q, Ye Z, Shen C, Li J, Zhang L, *et al*. Asymmetric subgenome selection and *cis*-regulatory divergence during cotton domestication. *Nat Genet*. 2017:**49**(4):579–587. https://doi.org/10.1038/ng.3807.

Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, Zhang X. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants*. 2018:**4**(2):90–97. https://doi.org/10.1038/s41477-017-0096-3.

Wang G, Zhou N, Chen Q, Yang Y, Yang Y, Duan Y. Gradual genome size evolution and polyploidy in *Allium* from the Qinghai-Tibetan Plateau. *Ann Bot*. 2021:**131**(1):109–122. https://doi.org/10.1093/aob/mcab155.

Weinberg-Shukron A, Ben-Yair R, Takahashi N, Dunjić M, Shtrikman A, Edwards CA, Ferguson-Smith AC, Stelzer Y. Balanced gene dosage control rather than parental origin underpins genomic imprinting. *Nat Commun*. 2022:**13**(1):4391. https://doi.org/10.1038/s41467-022-32144-z.

Weintraub H, Groudine M. Chromosomal subunits in active genes have an altered conformation. *Science*. 1976:**193**(4256):848–856. https://doi.org/10.1126/science.948749.

Wendel JF. The wondrous cycles of polyploidy in plants. *Am J Bot*. 2015:**102**(11):1753–1756. https://doi.org/10.3732/ajb.1500320.

Wendel JF, Albert VA. Phylogenetics of the cotton genus (*Gossypium*): character-state weighted parsimony analysis of chloroplast-DNA restriction site data and its systematic and biogeographic implications. *Syst Bot*. 1992:**17**(1):115. https://doi.org/10.2307/2419069.

Wendel JF, Brubaker CL, Seelanan T. The origin and evolution of *Gossypium*. In: Stewart JM, Oosterhuis DM, Heitholt JJ, Mauney JR, editors. *Physiology of cotton*. Netherlands: Springer; 2010. pp. 1–18.

Wendel JF, Cronn RC. Polyploidy and the evolutionary history of cotton. *Adv Agron*. 2003:**78**:139–186. https://doi.org/10.1016/S0065-2113(02)78004-8.

Wendel JF, Grover CE. Taxonomy and evolution of the cotton genus, *Gossypium*. In: Fang DD, Percy RG, editors. *Cotton*. 2nd ed. Madison (WI): American Society of Agronomy; 2015. p. 25–42.

Wendel JF, Lisch D, Hu G, Mason AS. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. *Curr Opin Genet Dev*. 2018:**49**:1–7. https://doi.org/10.1016/j.gde.2018.01.004.

Wendel JF, Olson PD, Stewart JM. Genetic diversity, introgression, and independent domestication of old world cultivated cottons. *Am J Bot*. 1989:**76**(12):1795–1806. https://doi.org/10.1002/j.1537-2197.1989.tb15169.x.

Wittkopp PJ, Haerum BK, Clark AG. Evolutionary changes in *cis* and *trans* gene regulation. *Nature*. 2004:**430**(6995):85–88. https://doi.org/10.1038/nature02698.

Wu C, Bingham PM, Livak KJ, Holmgren R, Elgin SC. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. *Cell*. 1979a:**16**(4):797–806. https://doi.org/10.1016/0092-8674(79)90095-3.

Wu C, Wong YC, Elgin SC. The chromatin structure of specific genes: II. Disruption of chromatin structure during gene activity. *Cell*. 1979b:**16**(4):807–814. https://doi.org/10.1016/0092-8674(79)90096-5.

Wu Y, Zhang W, Jiang J. Genome-wide nucleosome positioning is orchestrated by genomic regions associated with DNase I hypersensitivity in rice. *PLoS Genet*. 2014:**10**(5):e1004378. https://doi.org/10.1371/journal.pgen.1004378.

Xi L, Fondufe-Mittendorf Y, Xia L, Flatow J, Widom J, Wang J-P. Predicting nucleosome positioning using a duration Hidden Markov Model. *BMC Bioinformatics*. 2010:**11**(1):346. https://doi.org/10.1186/1471-2105-11-346.

Yoo MJ, Liu X, Pires JC, Soltis PS, Soltis DE. Nonadditive gene expression in polyploids. *Ann Rev Genet*. 2014:**48**(1):485–517. https://doi.org/10.1146/annurev-genet-120213-092159.

Yoo MJ, Szadkowski E, Wendel JF. Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity (Edinb)*. 2013:**110**(2):171–180. https://doi.org/10.1038/hdy.2012.94.

Yoo MJ, Wendel JF. Comparative evolutionary and developmental dynamics of the cotton (*Gossypium hirsutum*) fiber transcriptome. *PLoS Genet*. 2014:**10**(1):e1004073. https://doi.org/10.1371/journal.pgen.1004073.

You J, Lin M, Liu Z, Pei L, Long Y, Tu L, Zhang X, Wang M. Comparative genomic analyses reveal *cis*-regulatory divergence after polyploidization in cotton. *Crop J*. 2022:**10**(6):1545–1556. https://doi.org/10.1016/j.cj.2022.03.002.

Yu J, Jung S, Cheng C-H, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D. CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Res*. 2014:**42**(D1):D1229–D1236. https://doi.org/10.1093/nar/gkt1064.

Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015:**31**(14):2382–2383. https://doi.org/10.1093/bioinformatics/btv145.

Yuan J, Sun H, Wang Y, Li L, Chen S, Jiao W, Jia G, Wang L, Mao J, Ni Z, *et al*. Open chromatin interaction maps reveal functional regulatory elements and chromatin architecture variations during wheat evolution. *Genome Biol*. 2022:**23**(1):34. https://doi.org/10.1186/s13059-022-02611-3.

Zhang W, Jiang J. Application of MNase-seq in the global mapping of nucleosome positioning in plants. *Methods Mol Biol*. 2018:**1830**:353–366. https://doi.org/10.1007/978-1-4939-8657-6_21.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, *et al*. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008:**9**(9):R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhang K, Wang X, Cheng F. Plant polyploidy: origin, evolution, and its influence on crop domestication. *Hortic Plant J*. 2019:**5**(6):231–239. https://doi.org/10.1016/j.hpj.2019.11.003.

Zhang W, Wu Y, Schnable JC, Zeng Z, Freeling M, Crawford GE, Jiang J. High-resolution mapping of open chromatin in the rice genome. *Genome Res*. 2012:**22**(1):151–162. https://doi.org/10.1101/gr.131342.111.

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

Zhang K, Zhao X, Zhao Y, Zhang Z, Liu Z, Liu Z, Yu Y, Li J, Ma Y, Dong Y, *et al*. Cell type-specific cytonuclear co-evolution in three allopolyploid plant species. *Proc Natl Acad Sci U S A*. 2023:**120**(40):e2310881120. https://doi.org/10.1073/pnas.2310881120.

Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res*. 1998:**8**(5):479–492. https://doi.org/10.1101/gr.8.5.479.

Zhao L, Yang Y-Y, Qu X-J, Ma H, Hu Y, Li H-T, Yi T-S, Li D-Z. Phylotranscriptomic analyses reveal multiple whole-genome duplication events, the history of diversification and adaptations in the Araceae. *Ann Bot*. 2022:**131**(1):199–214. https://doi.org/10.1093/aob/mcac062.

Zhao H, Zhang W, Chen L, Wang L, Marand AP, Wu Y, Jiang J. Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol*. 2018:**176**(4):2789–2803. https://doi.org/10.1104/pp.17.01467.

Zhao H, Zhang W, Zhang T, Lin Y, Hu Y, Fang C, Jiang J. Genome-wide MNase hypersensitivity assay unveils distinct classes of open chromatin associated with H3K27me3 and DNA methylation in *Arabidopsis thaliana*. *Genome Biol*. 2020:**21**(1):24. https://doi.org/10.1186/s13059-020-1927-5.

Zheng D, Ye W, Song Q, Han F, Zhang T, Chen ZJ. Histone modifications define expression bias of homoeologous genomes in allotetraploid cotton. *Plant Physiol*. 2016:**172**(3):1760–1771. https://doi.org/10.1104/pp.16.01210.

Zlatanova J, Seebart C, Tomschik M. The linker-protein network: control of nucleosomal DNA accessibility. *Trends Biochem Sci*. 2008:**33**(6):247–253. https://doi.org/10.1016/j.tibs.2008.04.001.