PNAS RESEARCH ARTICLE | EVOLUTION

Genome-wide chromatin accessibility analysis unveils open chromatin convergent evolution during polyploidization in cotton

Jinlei Han[a,1], Damar Lopez-Arredondo[b,1], Guangrun Yu[c,1], Yankun Wang[c,1], Baohua Wang[a,1], Sarah Brooke Wall[d], Xin Zhang[a], Hui Fang[a], Alfonso Carlos Barragán-Rosillo[b], Xiaoping Pan[d], Yanqin Jiang[e], Jingbo Chen[e], Hui Zhang[a], Bao-Liang Zhou[f], Luis Herrera-Estrella[b,g,2], Baohong Zhang[d,2], and Kai Wang[a,2]

Contributed by Luis Herrera-Estrella; received June 6, 2022; accepted September 8, 2022; reviewed by Martin Lysak, Christopher Saski, and Yijing Zhang

Allopolyploidization, resulting in divergent genomes in the same cell, is believed to trigger a "genome shock", leading to broad genetic and epigenetic changes. However, little is understood about chromatin and gene-expression dynamics as underlying driving forces during allopolyploidization. Here, we examined the genome-wide DNase I-hypersensitive site (DHS) and its variations in domesticated allotetraploid cotton (*Gossypium hirsutum* and *Gossypium barbadense*, AADD) and its extant AA (*Gossypium arboreum*) and DD (*Gossypium raimondii*) progenitors. We observed distinct DHS distributions between *G. arboreum* and *G. raimondii*. In contrast, the DHSs of the two subgenomes of *G. hirsutum* and *G. barbadense* showed a convergent distribution. This convergent distribution of DHS was also present in the wild allotetraploids *Gossypium darwinii* and *G. hirsutum* var. *yucatanense*, but absent from a resynthesized hybrid of *G. arboreum* and *G. raimondii*, suggesting that it may be a common feature in polyploids, and not a consequence of domestication after polyploidization. We revealed that putative *cis*-regulatory elements (CREs) derived from polyploidization-related DHSs were dominated by several families, including Dof, ERF48, and BPC1. Strikingly, 56.6% of polyploidization-related DHSs were derived from transposable elements (TEs). Moreover, we observed positive correlations between DHS accessibility and the histone marks H3K4me3, H3K27me3, H3K36me3, H3K27ac, and H3K9ac, indicating that coordinated interplay among histone modifications, TEs, and CREs drives the DHS landscape dynamics under polyploidization. Collectively, these findings advance our understanding of the regulatory architecture in plants and underscore the complexity of regulome evolution during polyploidization.

polyploidization | chromatin accessibility | cotton | genome evolution | histone modification

Significance

Polyploidization may cause reshuffling of gene expression accompanied by changes in chromatin-structure dynamic. In eukaryotes, gene expression is regulated via the complex interplay between transcription factors and *cis*-regulatory DNA elements (CREs) in regions of open chromatin. We reveal that chromatin accessibility, displaying distinct distributions in diploid progenitors, shows a convergent change during polyploidization in cotton. We propose that this convergent change might be a common feature in polyploids, and not a consequence of domestication after polyploidization. Moreover, we show that the coordinated interplay among histone modifications, transposable elements, and CREs drove open chromatin dynamics under polyploidization. This study presents a holistic view of the regulatory landscape changes during polyploidization.

Polyploidization refers to an event whereby two or more genomes are brought together in the same nucleus and is a major force in plant evolution and speciation (1, 2). Moreover, many other plants containing duplicated chromosomes or chromosomal segments reflect ancient or recent rounds of polyploidy. Autopolyploidy and allopolyploidy are the two major types of polyploidy: The former arises from whole-genome duplication within a single species, and the latter usually forms through the combined processes of interspecific hybridization and genome doubling (3, 4). Allopolyploidization resulting in divergent genomes in a single cell is believed to trigger "genome shock" and to cause broad gene-expression aberrations, which could be detrimental to newly formed polyploidy (5, 6). Continuous efforts have revealed that epigenetic modification changes (7–9), homologous recombination (10), three-dimensional chromatin conformation (11, 12), transposable element (TE) reactivation (13), and small interfering RNA expression (14) are important factors driving gene-expression bias or novelty in plant polyploidization. However, there is still a lack of understanding of the mechanisms controlling gene transcription changes mediated by *cis*-regulatory elements (CREs) after plant polyploidization.

Genomic regions containing active CREs are open or accessible to regulatory proteins (i.e., open chromatin) because of the eviction or displacement of nucleosomes in local chromatin (15). A typical feature of open chromatin is increased sensitivity to nuclease (e.g., DNase I) digestion, creating DNase I-hypersensitive sites (DHSs), which are considered a hallmark of active regulatory DNA in eukaryotic genomes (16). Recent advances in methods of open chromatin analysis, including DHS identification combined with high-throughput sequencing (DNase-seq) and the assay for transposase-accessible chromatin sequencing (ATAC-seq), have provided detailed genome-wide information on CREs in diverse cell types, tissues, and developmental stages in both animals and plants

The authors declare no competing interest.

[1]J.H., D.L.-A., G.Y., Y.W., and B.W. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: luis.herrera-estrella@ttu.edu or zhangb@ecu.edu or kwang5@ntu.edu.cn.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2209743119/-/DCSupplemental.

Published October 24, 2022.

(17–22). In addition, histone modifications have been found to be tightly coordinate with open chromatin in plants. Open chromatins that overlap with gene or locate within 1 kb upstream of the transcription start site (TSS) are enriched with particular active histone modifications, including H3K4me3, H3K56ac, and H3K36me3 (21, 23). In contrast, distal open chromatin regions (>1 kb from gene) can be modified by H3K56ac or H3K27me3, which may act as an enhancer and repressor, respectively (21). Moreover, open chromatin in plant enhancers can be characterized by both active and inactive marks, including H3K27me3, H3K4me3, H3K27ac (24), and H3K9ac (25). Thus, the tight correlation between open chromatin and histone modification indicates indispensable roles for histone modifications in the functionating of open chromatin.

Cotton (*Gossypium*), a major source of natural textile fiber, is a particularly powerful model for obtaining genetic and epigenetic insights into polyploids. This genus consists of ∼50 species, including 40 to 45 diploids ($2n = 2x = 26$) and 7 allotetraploids ($2n = 4x = 52$) (26, 27). Two diploid species, *Gossypium arboreum* (Ga) and *Gossypium raimondii* (Gr), diverged from a common ancestor ∼4.7 million to 5.2 million years ago (Mya), and their genomes subsequently remerged via a polyploidization event that occurred 1 to 1.6 Mya (28, 29). Thereafter, diversification and domestication gave rise to seven divergent allotetraploids, such as cultivated *Gossypium hirsutum* (Gh) and *Gossypium barbadense* (Gb) and wild *Gossypium darwinii* (Gd) (Fig. 1*A*). Gh var. *yucatanense* (wGh) is considered as a wild type of Gh that gave rise to cultivated Gh by domestication (30). In tetraploid cotton, the expression of homeologous gene pairs from two subgenomes tends to be in a balanced pattern (i.e., similar levels of expression in both genomes) or slightly biased toward the A-genome (31), while synthetic allopolyploid from the putative ancestors Ga and Gr displayed unbalanced expression toward the D-genome (32). Although genetic and epigenetic factors, including genome architecture (33), DNA methylation (9), and the neofunctionalization of long noncoding RNA transcripts (34), are thought to be involved in the changes in gene expression during polyploidization, the role of chromatin remodeling on postpolyploid genome evolution and domestication is still largely unclear.

Here, we developed whole-genome DHS maps of these two diploid progenitors (Ga and Gr), F1 hybrid (Ga × Gr), domesticated (Gh and Gb), and wild (Gd and wGh) allotetraploid cotton by DNase-seq. Our results demonstrate a convergent change of DHS landscape during cotton polyploidization. We propose the existence of coordinated interplay among histone modification, TEs, and specific transcription factor (TF) families underlying the observed regulome changes. The results reveal unique insights into regulome evolution under plant polyploidization and provide a valuable resource regarding the *cis*-regulatory landscape and histone modifications of cotton for further research and agronomic trait improvement.

Results

Genome-Wide Identification of DHSs in Diploid and Tetraploid Cotton. To elucidate the dynamics of open chromatin landscape under polyploidization, we constructed three DNase-seq libraries (each with three replicates) using leaf tissues for the cultivated tetraploid cotton Gh cv. TM-1 and its extant diploid progenitors Ga and Gr (*Materials and Methods*). A total of 784 million, 596 million, and 717 million paired-end reads (150 bp) were obtained, corresponding to 102× coverage of the Gh genome (2.3 Gb) (35), 105× coverage of the Ga genome (1.7 Gb) (36), and 292× coverage of the Gr genome (0.7 Gb) (37, 38), respectively (Fig. 1*A* and *SI Appendix*, Table S1). By mapping to the respective reference genomes, we obtained ∼426 million, 327 million, and 381 million unique reads for Gh, Ga, and Gr, respectively. Spearman's rank correlation coefficients between biological replicates ranged from 0.84 to 0.91 (*SI Appendix*, Fig. S1), indicating that duplicates were highly correlated. DHSs were then identified, and only consensus DHSs that were detected in all three replicates, with a minimum of 1-bp overlap, were retained for downstream analyses. In total, 133,604 reproducible DHSs from Gh, 77,915 from Ga, and 59,997 from Gr were identified (Fig. 1 *B* and *C*), demonstrating a positive correlation between DHS number and genome size. The genome distribution of DHSs displayed a similar trend as gene density in all three species (*SI Appendix*, Fig. S2), which is consistent with previous studies in both plants and animals (16, 39). Additionally, DNase I accessibility was highly positively correlated with the expression of nearest-neighbor gene, and highly expressed genes contained longer DHSs around their TSSs ($P < 0.01$, Wilcoxon test) (Fig. 1*C*). Because our strategy was based on the double-hit approach (40), to further test the reproducibility of our data, we produced a Gr DNase-seq library by the end-capture strategy, which has been widely used in plants and animals (41). The high rank of data correlations (Spearman's rank correlation coefficients, 0.80 to 0.82) (Fig. 1D and *SI Appendix*, Fig. S3) between end-capture and double-hit strategies indicate that the two datasets were highly correlated and reproducible.

DHSs in the A and D Genomes Present Convergent Distributions after Polyploidization. For genomic distribution analysis, DHSs were divided into three categories, according to their distance to the nearest gene: genic (overlapped with a gene), proximal (within 1 kb upstream of the TSS or 1 kb downstream of the transcription termination site), and distal (>1 kb from any gene) (Fig. 2*A*). We observed distinct proportion of DHSs in the three categories between Ga and Gr, especially in the distal (56.4% in Ga vs. 36.8% in Gr) and genic (18.4% in Ga vs. 41.8% in Gr) regions (Fig. 2*A*). Since the sequencing depth for Gr is higher than that for Ga, it may have impacted DHS calling (20, 42) and the subsequent distribution assays. Thus, we conducted DHS calling from Gr by using datasets with the same Ga sequencing depth. When Ga and Gr were compared, differences in DHS proportions in distal and genic regions were still observed (*SI Appendix*, Fig. S4), confirming the distinct DHS distributions between these two genomes. Intriguingly, this was not observed between the A and D subgenomes of Gh, designated GhAT and GhDT, respectively, where T refers to the tetraploid nature of the entry (Fig. 2*A*). In contrast, we observed largely similar proportions of DHSs for all three categories between GhAT and GhDT, suggesting a convergent distribution of DHSs.

To further confirm the convergent distribution of DHSs, we conducted a DHS-number survey of homeologous tetrad genes. A total of 18,493 homeologous tetrad genes were identified; each tetrad gene showed a 1:1:1:1 correspondence across Ga, Gr, GhAT, and GhDT genomes (*Materials and Methods*). Among them, 14,015 (75.7% of 18,493) tetrads had at least one genic DHS (i.e., at least one genic DHS from four genes), and 4,478 tetrads had no DHS in any of the four genes in a given tetrad. The 14,015 tetrad genes were then analyzed to assess whether the A and D homeologs had equal DHS numbers (A = D) or not (A ≠ D) (Dataset S1). We found that the majority (9,502; 67.8% of 14,015) of homologous genes showed an A ≠ D pattern between Ga and Gr (Fig. 2*B*), which is consistent with the distinct DHS proportions (Fig. 2*A*).

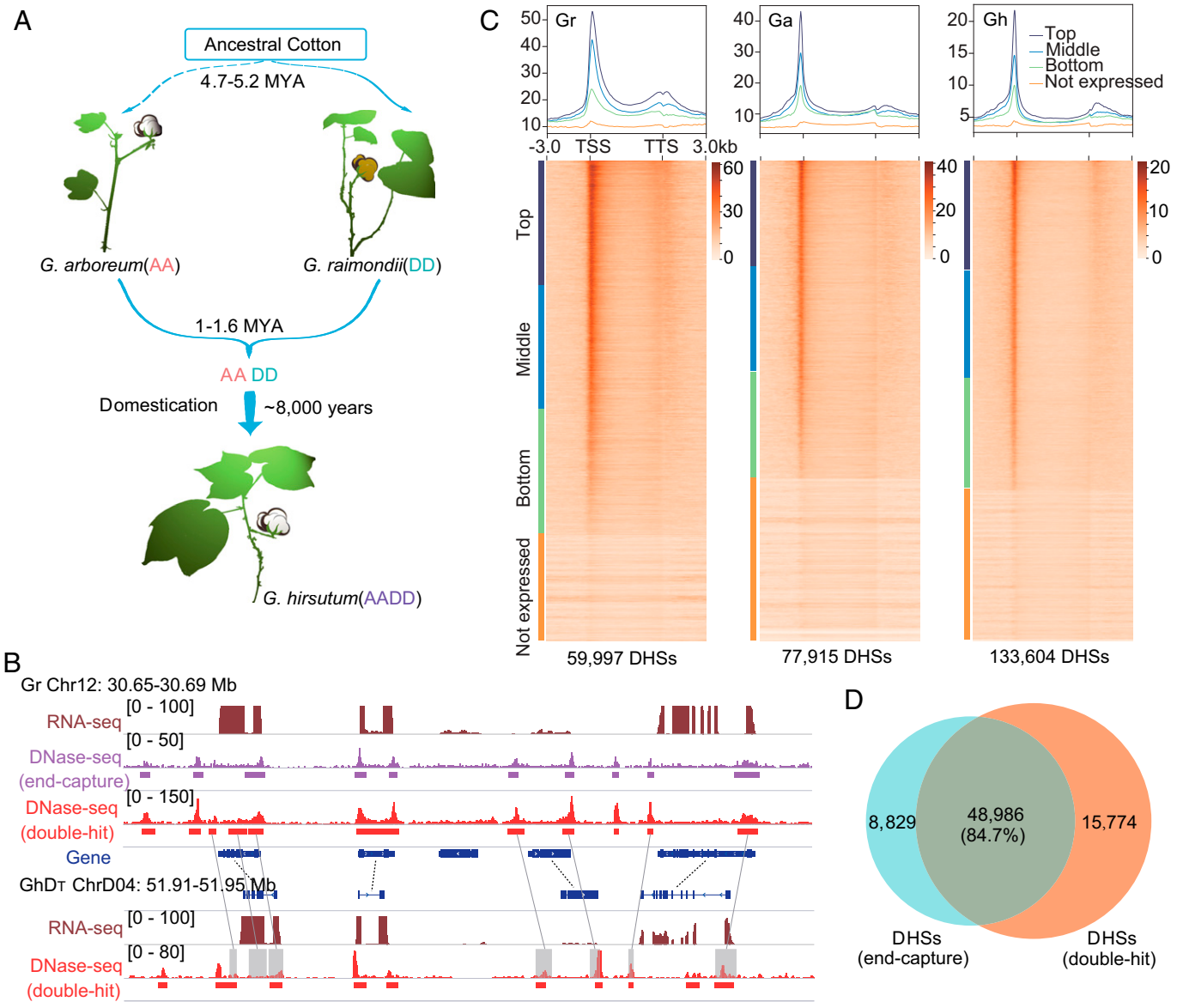

Fig. 1. Identification of DHSs in the cotton genome. (*A*) Schematic model of the evolutionary history of diploid and allotetraploid cotton. (*B*) Genomic tracks illustrating gene-expression (RNA-seq) and chromatin-accessibility (DNase-seq) profiles across a syntenic region between Gr and GhDT. Orthologous genes are connected by dotted lines. The DNA sequence of each Gr DHS was searched in GhDT, and the homologous region is indicated by gray shading. (*C*) Chromatin accessibility around genes in different cotton species. Genes are ordered by expression (highest to lowest). Expressed genes with FPKM values greater than 0.1 were equally divided into three groups based on their expression levels, ranging from high expression (Top) to low expression (Bottom). Metaplots above each heatmap are derived from genes binned by expression levels: top, middle, bottom, and not expressed (FPKM < 0.1). (*D*) Overlap assays between double-hit and end-capture DNase-seq peaks.

However, the proportion of genes with an A ≠ D pattern (4,336; 30.9%) was significantly lower ($P < 0.01$, Fisher's exact test) and higher A = D (9,679; 69.1%) between the two Gh subgenomes than between the Ga and Gr genomes. By analyzing individual gene pairs, we found that 75.2% of A ≠ D gene pairs (7,149 of 9,502) between Ga and Gr transitioned to the A = D pattern in Gh, which suggests a transition from the dominant pattern A ≠ D in Gr/Ga to A = D in Gh. Moreover, there were substantially fewer genes with high transcriptional differences between GhAT and GhDT than between Ga and Gr (Fig. 2*C* and Dataset S1). Gene Ontology (GO) enrichment analysis revealed that the genes with an A ≠ D to A = D transition pattern were enriched with the terms related to biological regulation, such as regulation of transcription, RNA metabolic and biosynthetic process, and nucleic acid-templated transcription (*SI Appendix*, Table S2). Therefore, the DHS distribution changes between Gh and their putative ancestors suggest that the open chromatin landscapes have been subjected to reprogramming through polyploidization and tend to present a convergent distribution between subgenomes of the tetraploid spcies. Moreover, the convergent changes in DHS distributions might be the result of convergent changes in gene expression during polyploidization in cotton.

The Convergent Distribution of DHSs Does Not Develop in the Early Stage of Polyploidization. To investigate the prevalence of convergent DHS distribution in tetraploid cotton, we conducted DNase-seq in another cultivated tetraploid cotton, Gb. We identified a total of 95,801 DHSs by generating 261 million DNase-seq reads (*SI Appendix*, Fig. S5 and Table S1). Convergent DHS distribution was also observed in the two subgenomes of Gb (Fig. 2*A*). Since the Gh (cv. TM-1) and Gb accessions we analyzed are both cultivars, we asked whether the convergent distribution of DHSs was a consequence of domestication. To explore

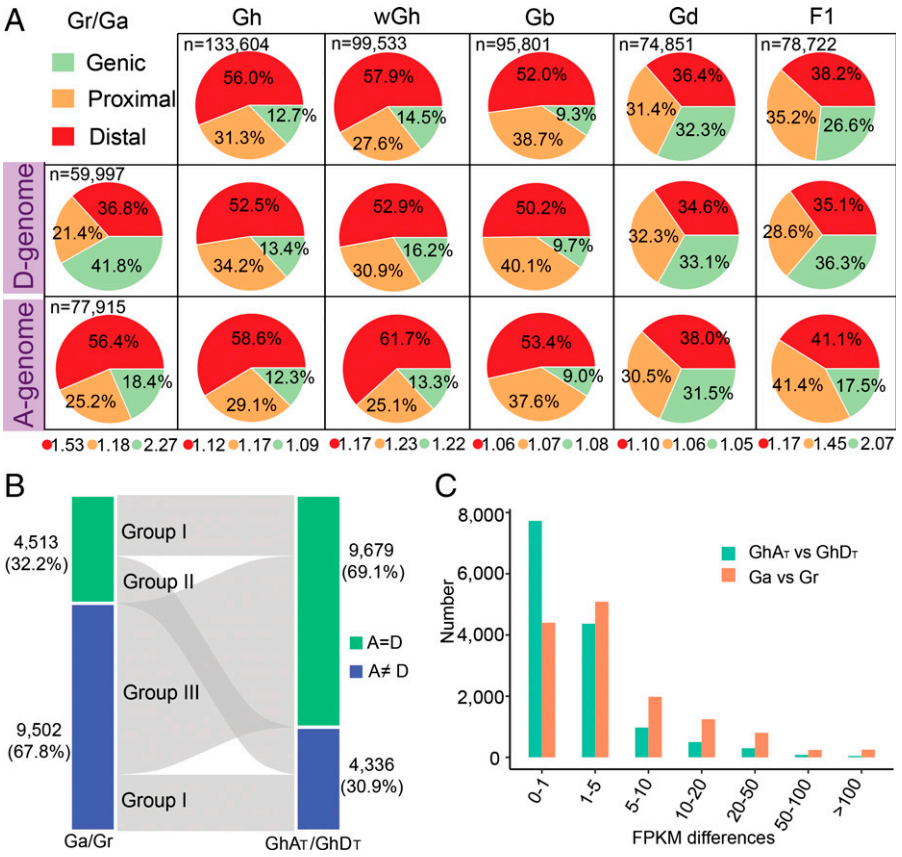

Fig. 2. Genomic distribution of DHSs in different cotton species. (*A*) The proportion of DHSs that are categorized as genic, proximal, and distal within different cotton species. The number of DHSs for each cotton accession is indicated. The fold change (high value/low value) of DHS proportions between the A genome and D genome in each genome region is given at the bottom. (*B*) DHS number comparison between homeologous genes. (*C*) Statistics of homeologous genes and expression differences. To compare the expression levels of homeologous genes, FPKM values were calculated. The gene number was calculated as the difference in FPKM value as follows: 0 to 1, 1 to 5, 5 to 10, 10 to 20, 20 to 50, 50 to 100, and >100.

this possibility, we conducted DNase-seq of wild tetraploid cotton wGh and Gd (Fig. 2*A*) and identified a total of 99,533 and 74,851 DHSs, respectively (*SI Appendix*, Fig. S5 and Table S1). Interestingly, we observed convergent DHS distributions between the two subgenomes in both wGh and Gd (Fig. 2*A*), indicating that the convergent distributions of open chromatin landscapes in tetraploid cotton are a consequence of polyploidization, rather than domestication.

Next, we analyzed Ga × Gr F1 plants to investigate whether convergence in DHS distribution between subgenomes arose immediately after hybridization. A total of 78,722 DHSs were identified in F1 plants (295 million DNase-seq reads) (*SI Appendix*, Fig. S5 and Table S1), showing clear divergent distribution of DHS between the A subgenome (F1A) and the D subgenome (F1D) (Fig. 2*A*). These results showed that DHS distribution convergence is not established immediately after hybridization. Interestingly, percentages of distal DHSs showed differences among cotton species with different genome size and polyploidy levels. For example, Ga, Gh, wGh, and Gb had 52.0 to 57.9% distal DHSs. By contrast, Gr, Gd, and Ga × Gr F1 contained 36.4 to 38.2% distal DHSs. This result underlines the fact that the proportion of distal DHS is not tightly correlated with genome size and polyploidy levels.

To assess the open chromatin landscape of the two parents without the impact of hybridization, an in silico hybrid was constructed by mixing the diploid parental DNase-seq data at a ratio of 1.71:0.76 (according to the genome size ratio of Ga/Gr) and down-sampling to the same sequencing depth as F1. We observed high levels of DHS overlap between F1 plants and the parents at the overall genome (83.7%) and subgenome (81.2% for A subgenome and 87.5% for D subgenome) levels (*SI Appendix*, Fig. S6), suggesting that the state of the open chromatin landscape was stable after interspecific hybridization. In addition, a total of 23,443 DHSs in F1 that differed from the diploid parents were identified (*SI Appendix*, Fig. S6). The nearest genes to these DHSs were enriched in GO categories such as regulation of gene expression, nitrogen compound metabolic process, and macromolecule biosynthetic process (*SI Appendix*, Table S3). Taken together, these results suggest that the convergent DHS distribution of tetraploids likely developed gradually, rather than occurring immediately after polyploidization.

Distinct Open Chromatin Reprogramming Accompanied by Different Steps of Allopolyploidization. DNase-seq data from diploid parents [Ga (AA) and Gr (DD)], the F1 AD hybrid, and tetraploid wGh (AADD) and Gh (AADD), allowed us to investigate reprogramming of the chromatin-accessibility response to different events during allopolyploidization, including interspecific hybridization (F1 vs. parents), genome doubling and evolution (wGh vs. F1), and domestication (Gh vs. wGh). To improve data comparability, we mapped each of the DNase-seq datasets to the same reference sequences of Gh (35), as previously done in wheat (43). Principal component analysis (PCA) of chromatin accessibility showed close clustering between F1 and its parents, Ga and Gr (Fig. 3*A*), suggesting slight changes in DHS landscapes caused by hybridization, consistent with their high levels of DHS overlap (>80%) (*SI Appendix*, Fig. S6). By contrast, Gh and wGh were

far away from the two diploids and F1, but closer together between them than with the diploids, suggesting that either polyploidization had a greater impact on chromatin remodeling than hybridization or domestication. In addition, 36.7% of A-genome and 48.1% of D-genome DHSs exhibited changes in accessibility (adjusted $P < 1e\text{-}5$) and demonstrated more significant accessibility changes for D genome than for the A genome ($P < 0.01$, Wilcoxon test) (Fig. 3*B*), indicating that these two genomes underwent alterations in open chromatin to different extents during allopolyploidization.

By comparing F1 and the two diploids (Ga and Gr), we discovered a total of 11,441 DHSs with significant accessibility increases in F1 (chromatin accessibility level fold change ≥ 2 and false discovery rate [FDR] ≤ 0.05), which represented hybridization-induced DHSs (hybDHSs) (Fig. 3 *C* and *D* and *SI Appendix*, Fig. S7). By using the same criteria, we identified 44,945 and 16,644 differential DHSs between wGh and F1 and Gh and wGh, respectively, by pairwise comparisons. These DHSs are thought to represent the genome-doubling and evolution-induced DHSs (deDHSs) and domestication-induced DHSs (domDHSs) (Fig. 3 *C* and *D* and *SI Appendix*, Fig. S7). The higher number of deDHSs than hybDHSs (3.9-fold) and domDHSs (2.7-fold) further supports the notion that genome doubling and subsequent long-term evolution caused a greater change in chromatin openness than hybridization and domestication. Notably, we observed a significant distal-region trend for the induced DHSs in all three groups ($P < 0.01$, Fisher's exact test), especially for deDHSs, as 84.0% of deDHSs were derived from the distal region (Fig. 3*E*). This result suggests that distal regulatory elements might be subject to a strong driving force during polyploidization in cotton.

DNA Motifs and TFs Associated with Differential DHSs. To further annotate the function of allopolyploidization-related differential DHSs, we searched for DNA-sequence motifs enriched in the hybDHSs, deDHSs, and domDHSs. A total of 342, 296, and 369 potential TF motifs were overrepresented (*E* value < 0.01) in hybDHSs, deDHSs, and domDHSs, respectively. Notably, 247 (55.0%) were present in all three DHS groups (Fig. 4*A*). Among these co-occurring motifs, the DNA binding with One Finger (Dof) family, which includes DOF4.7, DOF5.1, DOF5.6, and DOF5.8, were significantly enriched (Fig. 4 *B* and *C*); these motifs are known to control plant growth and the development of organs, including leaves, flowers, and vascular tissue (44–46). In addition, ERF48 and BPC1 binding motifs were also enriched in all three DHS groups (Fig. 4 *B* and *C*). The ERF48 motif consisted of a consecutive G sequence (Fig. 4*C*), which has been revealed as a target substrate of polycomb repressive complex 2 (PRC2) (47). The BPC1 motif resembles the canonical GAGA motif (Fig. 4*C*), which recruits GAGA-motif binding factors (GAFs). In both plants and animals, GAFs exhibit the same mechanistic function by recruiting the polycomb repressive complex (PRC) (48). Interestingly, PRCs, including PRC2, are histone methyltransferases that trimethylate H3K27 (H3K27me3), a mark of repressed chromatin in both plants and animals (49). This suggests that the repressive marker H3K27me3 may be involved in the open chromatin transitions during the process of cotton polyploidization.

We also observed 50, 22, and 66 motifs that occurred specifically for hybDHSs, deDHSs, and domDHSs, respectively (Fig. 4*A*). Interestingly, 70.0% (35/50) of hybDHS-specific motifs were associated with WRKY TFs (Fig. 4 *B* and *C*), which are mostly involved in the regulation of biotic and abiotic stress responses (50). A total of 54.5% (12/22) of deDHS-specific motifs were associated with NAC TFs (Fig. 4 *B* and *C*), which have been implicated in the development of plant architecture; for example, CUC1 and NAC45 are known to have key roles in shoot meristem formation (51) and root development (52). The identification of diverse dominant CREs suggests that these two evolutionary stages have specific impacts on the evolution of polyploid cotton. Among the 66 domDHS-specific motifs, GATA (10.6%, 7/66), HSF (10.6%, 7/66), and SBP (10.6%, 7/66) TFs were the most enriched (Fig. 4 *B* and *C*); these motifs have been reported to be involved in diverse functions, such as light (e.g., GATA19) and stress responses (e.g., HSFB2B) and vegetative to reproductive phase transition (e.g., SPL15) (53–55). These results suggest that the diverse TF binding sites found in domDHS-specific motifs may reflect artificial selection for the improvement of multiple traits in cultivated upland cotton during domestication.

Open Chromatin Alterations Coordinate with Epigenetic Modifications. Given that the interplay of DHSs and histone modifications has been well studied (21, 56), their cross-talk during long-term evolution after polyploidization is still unclear. Thus, we produced genome-wide maps of five histone modifications (H3K4me3, H3K27me3, H3K36me3, H3K27ac, and H3K9ac) in Ga, Gr, and Gh using chromatin immunoprecipitation and sequencing (ChIP-seq) (*SI Appendix*, Table S4 and Fig. S8). Chromatin modification features were similar among the three cotton species and other plants (23, 57–59). For example, the active histone modifications H3K4me3, H3K27ac, H3K9ac, and H3K36me3 were enriched in the TSS of active genes. In contrast, the repressive mark H3K27me3 was enriched in inactive genes across the gene body (*SI Appendix*, Fig. S9). Intriguingly, we observed an overall increase in signal enrichment at DHSs for the five histone marks (Fig. 5*A* and *SI Appendix*, Fig. S11; all DHSs in Ga, Gr, GhAT, and GhDT). These observations are contradictory to findings reported for other plants (21, 23, 25, 59–62), suggesting that open chromatin in cotton may adopt distinct nucleosomal modifications from other plants.

We analyzed differential DHSs between Ga and GhAT or Gr and GhDT by comparing chromatin accessibility (*Materials and Methods*) to assess the overall changes in open chromatin during polyploid formation and evolution. By integrating with ChIP-seq data of histone marks, we observed positive correlations between chromatin accessibility and the ChIP signals (Pearson's *r* of 0.460 to 0.781) (*SI Appendix*, Fig. S10). Moreover, ~20% of DHSs that did not show chromatin-accessibility differences had histone-modification changes (at least one mark) by comparing between Gh and Ga/Gr. However, when considering polyploidization-induced/repressed DHSs, the proportion of DHSs with differential histone modifications increased significantly to 43 to 71% ($P < 0.01$, Fisher's exact test). Thus, these results suggest that histone modifications played an important role in chromatin-accessibility dynamics during cotton polyploidization or domestication. Consistent with a positive correlation, polyploidization-induced DHSs (DHSs that occurred in Gh, accessibility fold change Gh/[Ga or Gr)] ≥ 2 and FDR ≤ 0.05) displayed markedly increased signals for all five marks compared with their counterparts in Ga or Gr (Fig. 5*A* and *SI Appendix*, Fig. S11). Conversely, polyploidization-repressed DHSs (DHSs that disappeared in Gh, accessibility fold change [Ga or Gr]/Gh ≥ 2 and FDR ≤ 0.05) exhibited obviously decreased signals for all five marks (Fig. 5*B* and *SI Appendix*, Fig. S11). We noted that the inactive mark H3K27me3, which indicated gene inactivation or heterochromatin, also displayed a positive correlation with accessibility change. This is in line with the observation that allopolyploidization-related differential DHSs

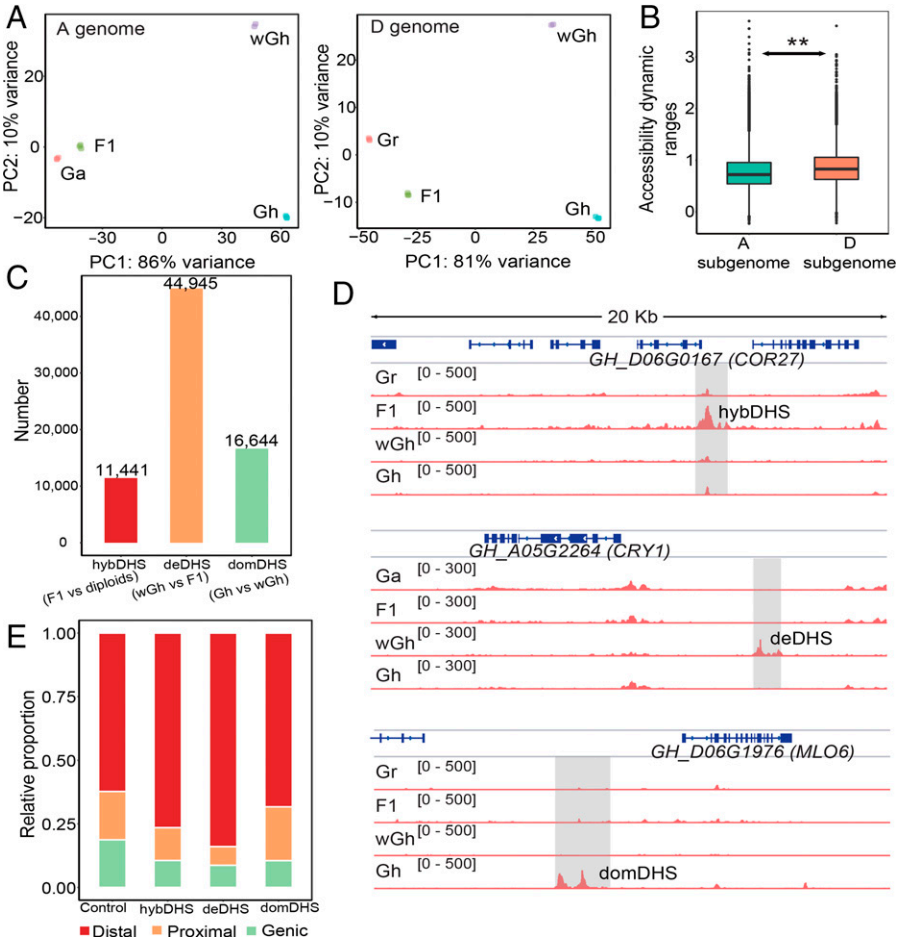

Fig. 3. Comparison of DHSs in diploid progenitors and their descendent genomes in tetraploid cotton. (*A*) PCA plots of DNase-seq data. (*B*) Comparison of accessibility variation of DHSs between the A subgenome and the D subgenome. The box plot shows the distributions of the dynamic ranges [log10(max-min)] of accessibility levels for DHSs in different subgenomes. The accessibility level was measured by quantifying the DNase-seq reads within each DHS and normalizing by per million mapped reads. The Wilcoxon test was used to analyze significance. $^{**}P < 0.01$. (*C*) The numbers of hybDHSs, deDHSs, and domDHSs. (*D*) Representative examples of hybDHSs (*Top*), deDHSs (*Middle*), and domDHSs (*Bottom*). The locations of the DHSs are shaded. (*E*) The proportions of genic, proximal, and distal DHSs within each group of hybDHSs, deDHSs, and domDHSs. The results calculated from all DHSs were used as the control.

were enriched in DNA motifs targeted directly or indirectly by PRCs (function in catalyzing the trimethylation of H3K27) (Fig. 4*B*). Thus, these results indicate that both active and inactive epigenetic marks are correlated with the regulation of chromatin accessibility during the long-term process of polyploidization in cotton.

TEs Contribute to DHS Formation in Polyploidization. Given the important role of TEs on the origin and diversity of CREs (60, 63), we wondered to what extent TEs impacted DHS dynamics related to polyploidization. We identified a total of 16,353, 3,699, and 26,129 DHSs associated with TEs (*Materials and Methods*), accounting for 21.0%, 6.2%, and 19.6% of total DHSs identified in the genomes of Ga, Gr, and Gh, respectively (*SI Appendix*, Fig. S12). The substantial number of TE-associated DHSs suggest a potential contribution of TEs to DHS formation in each species of cotton.

Then, we identified the "newborn DHSs" (nbDHSs), open chromatin regions present only in Gh (not present in Ga or Gr) that could be derived from TE-related sequence amplification during or after polyploidization. A total of 2,320 nbDHSs (1,033 in GhAT and 1,287 in GhDT) were identified. Strikingly, 1,314 (56.6%) of these 2,320 nbDHSs were defined as TE-associated DHSs (referred to as TE-nbDHSs) (Fig. 6*A*). This is likely an underestimation because TEs are highly repetitive, and most of them may have been discarded in our unique read-based analysis. These results indicate that TEs contribute profoundly to reshape chromatin architecture and gene expression during cotton allopolyploidization.

Gypsy-type retrotransposons played a dominant role in the TE-nbDHSs (616 of TE-derived nbDHSs; 46.9%) (*SI Appendix*, Fig. S13). This result is in line with the proportion of Gypsy elements among TEs in the Gh genome (45.7%). DNA transposons of hAT and CACTA elements accounted for 5.6% (73) and 7.8% (103) of the TE-nbDHSs, respectively (*SI Appendix*, Fig. S13). In contrast, the hAT and CACTA elements accounted for only 1.6% and 4.2% of all TEs in the Gh genome, respectively, which was significantly lower than the percentages accounted for by TE-nbDHSs ($P < 0.01$, Fisher's exact test). This may suggest that hAT and CACTA elements have distinct impacts on nbDHS formation than other types of TEs during cotton allopolyploidization.

TE-Associated DHSs Act as Functional Regulatory Elements. We found that the majority (98.0%; 1,288 of 1,314) of nbDHS-associated TEs were silent (reads per kilobase of exon per million reads mapped [RPKM] ≤ 1, accounting for 82.9%) or had very few transcripts ($0 < \text{RPKM} \leq 1$, 13%) (*SI Appendix*, Fig. S14), suggesting that most TE-nbDHSs do not act as CREs to regulate the expression of TEs themselves. Expression levels of genes

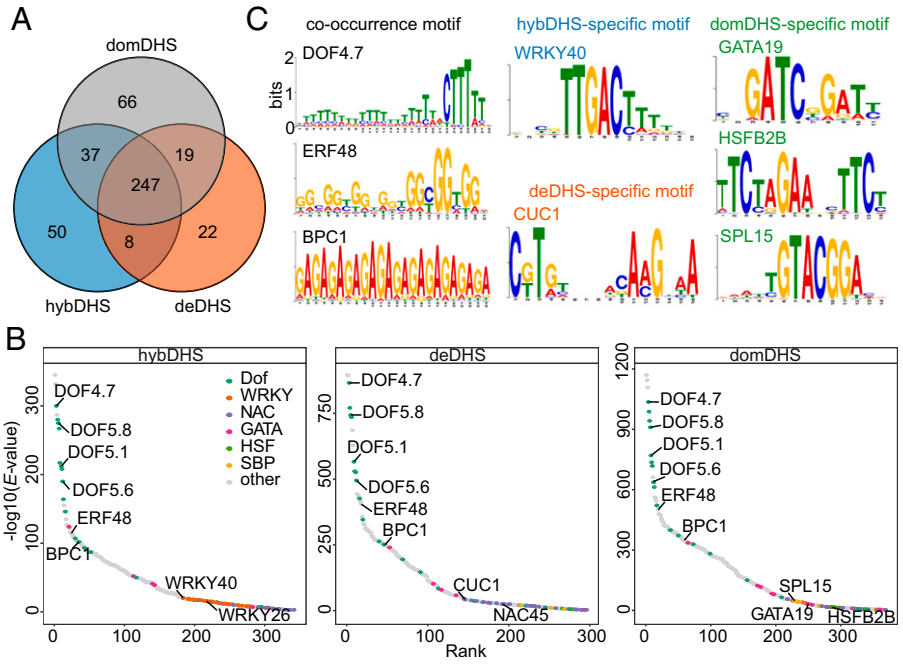

Fig. 4. Enrichment analysis of DHS-derived TF motifs. (*A*) Venn diagram showing the overlap of TF motifs identified in hybDHSs, deDHSs, and domDHSs. (*B*) Ranking of motifs enriched in hybDHSs, deDHSs, and domDHSs. TFs are indicated by colored dots. The *E* value of each motif was estimated with AME software. (*C*) Diagrams showing examples of motifs identified in a specific or all groups of hybDHSs, deDHSs, and domDHSs.

adjacent to TE-nbDHSs (1,136 genes) were higher than those of genes without DHSs ($P < 0.01$, Wilcoxon test) (Fig. 6*B*), indicating that TE-nbDHSs likely had a positive effect on the expression of adjacent genes. However, genes associated with TE-nbDHSs showed lower expression levels than genes associated with non-TE–DHSs ($P < 0.01$, Wilcoxon test) (Fig. 6*B*). Intriguingly, we found that only 38.2% (434/1,136) of genes adjacent to TE-nbDHSs had a homolog copy in Ga or Gr (Fig. 6*C*). Among these genes, 308 showed significant changes in expression, in contrast to their homologs in the Ga or Gr genome (expression fold changes ≥ 1.5) (Dataset S2). Thus, a total of 1,010 of the 1,136 (88.9%) TE-nbDHSs–associated genes showed expression bias or occurred specifically in Gh, indicating a highly dynamic state for genes potentially targeted by TE-nbDHSs.

We speculated that the functions of TE-nbDHS–associated genes are related to certain functions acquired by Gh through polyploidization. GO enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of these genes showed that they were highly enriched in response to stimulus, signal transduction, starch and sucrose metabolism, and circadian rhythm (Fig. 6*D*), which may explain some physiological changes in Gh compared to Ga or Gr, such as photoperiod insensitivity and an increase in fiber length. We then searched the TE-nbDHSs and identified a total of 64 enriched TF-binding DNA motifs (*E* value < 0.01) (Fig. 6*E*). We found that putative binding sites for TCP (18/64, 28.1%), ERF (12/64, 18.8%), and bHLH (11/64, 17.2%) TF family members were significantly enriched in TE-nbDHSs. These TFs are mainly involved in stress response, hormone signaling transduction, carbohydrate metabolism, and circadian clock regulation (64–66), which is consistent with the potential functions of TE-nbDHS–associated genes.

The identification of TE-nbDHSs could aid in the prediction regulators of their potential target genes. A sample is shown in Fig. 6*F*. *GH_D13G2387* is a cotton homolog of the *Arabidopsis* gene *CONSTANS-like 5*, which plays a role in the photoperiod and circadian clock pathway (67). We observed that the expression levels of *GH_D13G2387* were significantly higher (fragments per kilobase of transcript per million mapped reads [FPKM] value ratio = 3.8) than those of its ortholog in Gr (Fig. 6*F*). Two

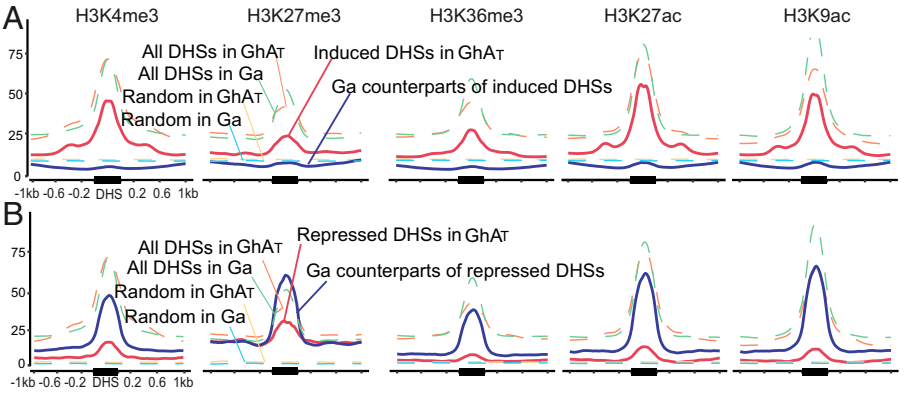

Fig. 5. The dynamics of DHS-associated histone modification during polyploidization in cotton. Histone modification of polyploidization-induced (*A*) or polyploidization-repressed (*B*) DHSs. Normalized ChIP-seq signals (RPKM) are shown (*y* axis).

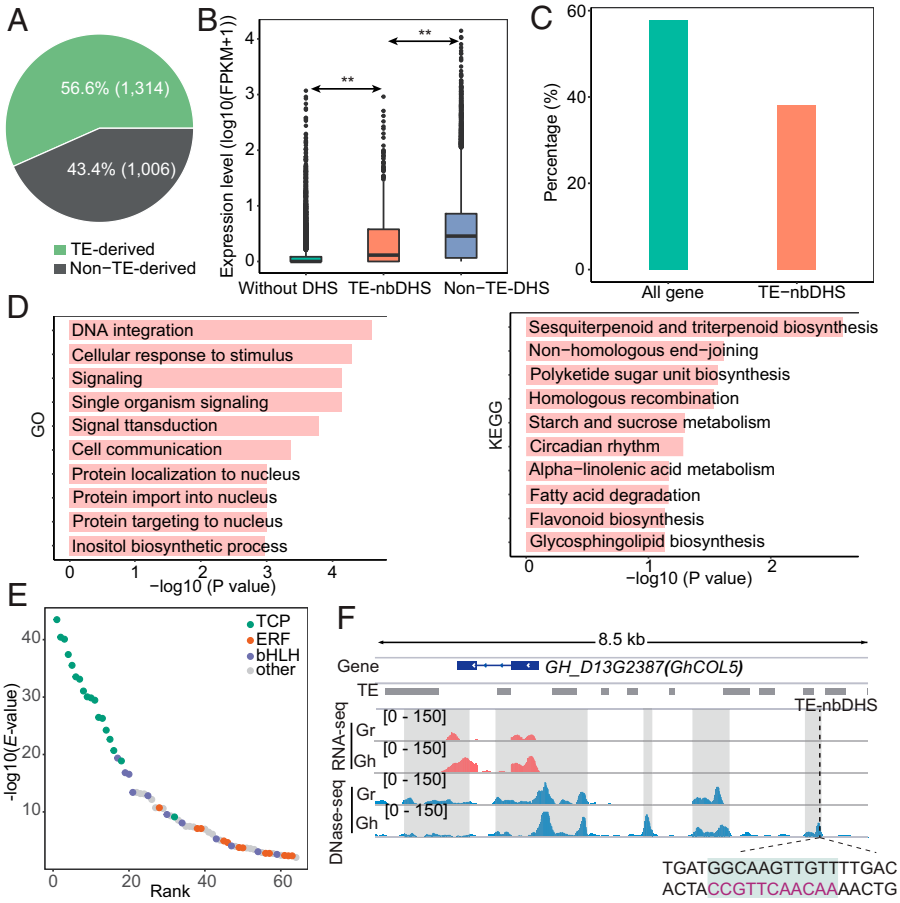

Fig. 6. The correlation of TEs and DHSs in cotton. (*A*) Proportion of TE-derived DHSs in relation to nbDHSs. (*B*) Expression levels of genes associated with different types of DHSs. "Without DHS" represents genes without DHSs. "TE-nbDHS" represents genes associated with TE-derived nbDHSs. "Non-TE-DHS" represents genes associated with only non-TE–DHSs. The Wilcoxon test was used to analyze significance. $^{**}P < 0.01$. (*C*) Proportion of genes with a 1:1 correspondence between Ga and GhAT or between Gr and GhDT. "TE-nbDHS" represents genes associated with TE-derived nbDHSs. (*D*) GO enrichment and KEGG pathways analyses of genes associated with TE-nbDHSs. The top 10 GO biological processes terms and KEGG pathways are indicated. (*E*) Ranking of motifs enriched in TE-nbDHSs. The colored spots represent different TF families. The *E* value of each motif was estimated with AME software. (*F*) A representative example, i.e., *GH_D13G2387*, a gene that shows up-regulated expression in Gh and is associated with TE-nbDHSs. The bHLH recognition sequence AACAACTTGCC (purple) was identified in the upstream TE-nbDHS. The locations of DHSs are shaded.

Gh-specific DHSs were found upstream of *GH_D13G2387*. Among them, one was identified as a TE-nbDHS located 5 kb upstream of *GH_D13G2387*. Motif scanning revealed that the TE-nbDHS contains the sequence AACAACTTGCC and is potentially targeted by bHLH TFs. In plants, several bHLH TFs act as transcriptional activators that can elevate the expression of *CONSTANS* genes (68). Therefore, we propose that bHLH element-associated TE-nbDHSs play a role in up-regulating the expression of *GH_D13G2387*, which may, in turn, contribute to the loss of photoperiod sensitivity in Gh.

Discussion

It is recognized that open chromatin reorganization occurs during major developmental phase transitions in plants (69, 70) and animals (71, 72), but open chromatin dynamics during genome polyploidization, as a common and crucial event in the evolution of flowering plants, is still poorly understood. Here, we report that the DHS landscape, which was different between two diploid ancestors, showed convergent distributions in domesticated tetraploid cotton. As DHSs are representative regulators of gene expression, the convergent distribution of DHSs in tetraploid cotton may provide insights into the interpretation of unbiased gene expression between the two subgenomes found in our (Fig. 2*C*) and previous (31, 73–75) studies. Furthermore, this convergent change in DHSs is proposed to not be a consequence of domestication after polyploidization, as the same distribution pattern was found in the wild tetraploid species Gd and wGh (Fig. 2*A*). This, in turn, indicates that the convergent DHS dynamics may be largely attributed to stabilize the genetic perturbations caused by polyploidization. Thus, determining whether the convergent evolution of open chromatin is a common feature of polyploidization is an interesting research focus because it may provide new insights into the evolutionary dynamics of CREs through plant evolution.

The lack of convergent DHS distribution in F1 plants indicates the gradualness of its formation (Fig. 2*A*). In fact, *cis*-regulatory divergence can act as a major driving force of gene evolution by subfunctionalization or neofunctionalization or be important for the expression of loci involved in quantitative traits (76, 77). However, evolution of the *cis*-regulatory repertoire has been found to occur over time (78) as a consequence of purifying selection against *trans*-regulatory diversity and positive selection for *cis*-regulatory changes (79, 80). Thus, this may explain the gradual formation of the convergent DHS distribution in cotton.

Similar to DHS landscape changes, DNA methylation has been found to show convergent and concerted changes between two subgenomes associated with polyploidization events in both allotetraploids cotton and *Arabidopsis* (8, 9). Among these species, two genomes with different methylation levels have been found to

show similar methylation levels after they merge in a single cell via polyploidization. More intriguingly, DNA methylation displays gradual changes as the convergent DHS distribution is established (8). The coincidental changes in DHSs and DNA methylation may suggest a correlation between epigenomic and regulatory landscape dynamics after plant polyploidization. In fact, DNA methylation and chromatin accessibility show strong correlations (DNA methylation is negatively correlated with chromatin accessibility) (81–83). Thus, we hypothesize that causal changes in open chromatin and histone modification are responsible, at least in part, for genome stabilization and adaptation during polyploidization (84).

Although the underlying force driving the convergent dynamics of the DHS landscape along with polyploidization is still unknown, our results uncovered several features of DHSs related to polyploidization in cotton. First, we observed that several TF binding motifs were dominant in polyploidization-related DHSs, including Dof, ERF48, and BPC1. Dof family TFs are known to be involved in the control of plant growth and the development of organs, including leaves, flowers, and vascular tissue (44–46). Thus, the enrichment of Dof motifs in polyploidization-related DHSs may reflect the typical phenotypic changes of increased growth and leaf or flower size in polyploids compared to diploids (29, 85). Interestingly, both ERF48 and BPC1 are involved in the recruitment of the histone methyltransferase H3K27me3 (47, 48). This is in line with the observation that allopolyploidization-induced DHSs were enriched in H3K27me3 signals (Fig. 5 and *SI Appendix*, Fig. S11) and highlights the effects of the repressive mark H3K27me3 on open chromatin dynamics during polyploidization in cotton.

Second, TEs acts as important contributors to the development of *cis*-regulatory sequences through their activity (86). The host genome can co-opt TEs into de novo regulatory elements (87). For example, in maize, a Copia-like TE was inserted ∼60 kb upstream of the *teosinte branched 1* gene to create an enhancer element that increased maize apical dominance (88). Examination of the whole genome showed that ∼25% DHSs were in TEs in maize (60). Here, we revealed that in polyploid cotton, nearly 20% of DHSs were derived from TEs. However, the proportion of TE-derived DHSs increased to 56.6% when newly occurring DHSs in polyploid cotton were examined (Fig. 6*A*). Furthermore, we revealed that most DHS-associated TEs were transcriptionally inactive, but remained functional with an open chromatin status. Therefore, these results suggest that TEs contribute substantially to the formation of DHSs related to polyploidization. Our analysis may have significantly underestimated the proportion of TE-nbDHSs related to polyploidization because of the repetitiveness of TE-related DHSs and the limited samples used in the current study (only young leaves).

Third, we revealed that all five histone-modification marks studied were enriched on cotton DHSs (Fig. 5 and *SI Appendix*, Fig. S11). Moreover, DHSs in response to polyploidization showed positive correlations with the signal intensities of the studied histone marks (Fig. 5 and *SI Appendix*, Fig. S11), suggesting that epigenetic modifications may be another factor accompanying or contributing to the convergent changes in DHSs after polyploidization. This hypothesis is supported by several studies in plants. For example, a study in rice showed that induced chromatin accessibility leads to preferential recruitment of H3K36me2 and H3K36me3 (89). In *Arabidopsis*, genome duplication can contribute to the switching of some loose and compact structural domains and altered H3K4me3 and H3K27me3 modification (90). We noted that the signals of the studied histone marks increased in the center of DHSs (Fig. 5 and *SI Appendix*, Fig. S11). This pattern is different from the findings in some other plant species, in which histone modifications are usually enriched at flanking regions of open chromatin and depleted from the center (21, 23, 25, 59–62). However, a similar pattern around DHSs can be found in human K562 cells (91, 92). If the difference in data treatment is ignored, this distinct pattern may suggest that open chromatin in cotton underwent dynamic nucleosomal modifications or displacement (23) and is distinct from that of other plants, which feature nucleosome-free DHSs. Therefore, whether the roles of histone modification in polyploidization events in cotton are applicable to other plants remains to be determined.

Materials and Methods

Plant Materials. All cotton materials, including Gr, Ga, interspecies hybrid (Ga × Gr) F1, Gh, Gb, Gd, and wGh, used in this study were grown under environment-controlled greenhouse conditions set to 13 h/11 h of light/dark, 28 °C light/26 °C dark, 60% humidity, and 270 μmol/m^2/s light intensity (cool-white fluorescent bulbs). Young leaf was collected and frozen immediately in liquid nitrogen for the subsequent experiments. In order to ensure that the leaves from different cotton lines were at the same developmental stage, the third and fourth true leaves were harvested (when the fifth true leaf emerged) for DNase-seq, RNA-seq, and ChIP-seq library constructions.

Library Construction for DNase-Seq. For the "double-hit" strategy, nuclei isolation, DNase I digestion, and DNase-seq library construction were performed per published protocol (93). Briefly, about 1 g of finely ground powder was suspended in the same volume of prechilled nuclear isolation buffer (10 mM Tris·HCl, 80 mM KCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 1 mM spermidine, 1 mM spermine, 0.15% mercaptoethanol, and 0.5 M sucrose, pH 9.5) as the volume of powder for nuclei isolation. Extracted nuclei were resuspended in nuclear digestion buffer (10 mM Tris·HCl, 10 mM NaCl, and 3 mM $MgCl_2$, pH 7.5) and were digested with gradient concentrations of DNase I for 10 min at 37 °C. DNA fragments of size < 250 bp from the optical DNase I treatment were isolated for library construction by using the NEBNext Ultra DNA Library Prep Kit (NEB, catalog [Cat.] no. E7370). The libraries were developed from three biological replicates for each cotton and were sequenced by using the Illumina HiSeq platform with a 150-bp pair-end model.

The "end-capture" strategy, in which a 20-bp fragment is extracted from the DNase I-digested chromatin end, was used to isolate the DHSs. DNase-seq experiments were performed exactly as described (20). Three biological replicates of DNase-seq libraries were developed from Gr. The target DNA fragments ∼90 bp had been purified for sequencing through an Illumina HiSeq platform with a 50-bp single-end model.

DNase-Seq Data Analysis. The raw reads generated from DNase-seq were quality-filtered and trimmed by using trim_galore v.0.6.4_dev (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Cleaned reads were mapped to their respective genome by using Bowtie2 version (v.) 2.3.4.1 (94) with default parameters. The genome sequence and annotation files for Gr (JGI-v2.1), Ga (CRI-v1.0), Gh (ZJU-v2.1), Gb (ZJU-v1.1), and Gd (HGS-v1.1) were downloaded from CottonGen (https://www.cottongen.org/). The F1 genome was generated by mixing the genome sequences of Ga and Gr. Mapped reads were then filtered by using SAMtools v.1.9 (95) to retain only correctly read pairs with a mapping-quality score of 10 or higher. Reads were further filtered to remove those mapping to either the chloroplast or mitochondrial genomes prior to further analysis. DHSs were identified by using the MACS2 v.2.1.4 (96) with the parameters "-f BAMPE –broad –nomodel –keep-dup all". For end-capture strategy, clean reads were mapped to the reference genome and processed to keep those with a mapping quality score of 10 or higher. DHSs were called by using MACS2 with the same parameters as mentioned above, except "-f BAM". Only consensus DHSs detected in all three replicates (with a minimum of 1-bp overlap) were retained for downstream analyses for both double-hit and end-capture experiments. Annotation of the DHSs relative to genes was performed with the annotatePeaks function of the HOMER v.4.11 package.

The allopolyploidization-related DHSs were identified as described (20, 43). Reads from Ga, Gr, F1, and wGh were mapped to the reference sequence of Gh.

The mapping and peak calling steps were performed as described above. DHSs from all species were merged to create a union set of DHSs. The read number for each species in the union set of DHSs was determined by using BEDTools v.2.29.2 (97). Three replicates from each species were counted, and the counts were processed by using DESeq2 (98). The PCA plot was generated by the plotPCA function of DESeq2. DHSs with a fold change ≥ 2 and FDR ≤ 0.05 by pairwise comparison were identified as allopolyploidization-related DHSs.

RNA-Seq and Data Analysis. For Gr, Ga, and Gh, total RNA of leaves from each replicate was extracted by using an Omega Plant RNA kit (Omega Bio-tek, Cat. no. R6827-01). RNA-seq libraries were prepared by using the Illumina TruSeq RNA Kit (NEB, Cat. no. E7530) and were sequenced by using an Illumina HiSeq system to produce 150-bp paired-end reads (*SI Appendix*, Table S5). The clean sequencing reads were mapped against the respective genome by using Tophat2 v.2.1.1 (99) with default settings. The Cufflinks v.2.2.1 (100) program was employed to calculate the normalized expression level (FPKM) of annotated genes. GO enrichment analysis was performed by using an online resource (https://www.omicshare.com/tools) with the default instructions.

ChIP-Seq and Data Analysis. ChIP-seq assay was performed by following a published protocol (101). Five commercial antibodies against H3K4me3 (Sigma, Cat. no. 07-473), H3K27me3 (Sigma, Cat. no. 07-449), H3K36me3 (Abcam, Cat. no. ab9050), H3K27ac (Sigma, Cat. no. 07-360), and H3K9ac (Sigma, Cat. no. 06-942) were used in immunoprecipitation. The Mock DNA was purified as a control. The ChIP and Mock DNAs were ligated with Illumina sequencing adaptors, size selection, and PCR amplification and sequenced with the Illumina HiSeq platform to produce 150-bp paired-end reads. ChIP-seq data from each species were obtained for two biological replicates. For each species, the data processing, including reads cleaning and mapping steps, was the same as described above for DNase-seq. Only correctly mapped read pairs with a mapping quality score of 10 or higher were used for further analysis.

To improve data comparability, reads from Ga and Gr were mapped to the reference sequence of Gh. The mapping and filtering steps were performed as described above. DHSs from Ga, Gr, and Gh were merged to create a union set of DHSs. The ChIP-seq read number for each species in the union set of DHSs was determined by using BEDTools v.2.29.2 (97). Two replicates from each species were counted, and the counts were processed by using DESeq2 (98). DHSs with a fold change ≥ 2 and FDR ≤ 0.05 between Ga and the A subgenome of Gh or between Gr and the D subgenome of Gh were identified as DHSs with differential histone-modification levels.

Data Visualization. For visualization, the filtered, sorted, and indexed BAM files were converted to the bigwig format by using the bamCoverage function in deepTools v.3.1.3 (102) with a bin size of 10 bp and RPKM normalization. Heatmaps and average plots displaying DNase-seq and ChIP-seq data were also generated by using the computeMatrix, plotHeatmap, and plotProfile functions in the deepTools package. Genome browser images were made by using the Integrative Genomics Viewer v.2.3.92 with bigwig files processed as described above.

Identification of Homeologous Tetrad Gene. The OrthoFinder v.2.3.8 (103) program was used to analyze the proteome of Gr, Ga, and the two subgenomes of Gh. One-to-one orthologous gene sets were then extracted from the results. For each tetrad gene, one, and only one, homeologous gene can be identified from each of the Ga, Gr, GhAT, and GhDT genomes.

Motif Analysis. We downloaded TF motif models from PlantTFDB (planttfdb.gao-lab.org/). The enriched motifs in the region of interest were identified by using the AME v.5.4.1 tool of MEME Suite (104) with the parameter settings "–control –shuffle– –scoring avg –method fisher". Motifs with an E value < 0.01 were retained. Potential TF binding sites were determined by scanning motif occurrences in the region of interest by using the FIMO v.5.4.1 tool of MEME Suite with the default settings.

Identification of TE-Derived DHS. TEs were identified in the Gh reference genomes by using the EDTA v.1.9.6 pipeline (105) with the parameters "–sensitive 1 –anno 1". The annotation from EDTA was then parsed and classified hierarchically into class 1 retrotransposons (including Gypsy, Copia, LINE, etc.), Class 2 DNA transposons (including CACTA, Mariner, Mutator, etc.), other, and unclassified categories. TE-derived DHSs were identified by determining the genomic coordinates of DHSs overlapping by at least 50% using the intersectBed program from BEDTools v.2.29.2 (97). Random control regions with the same length distribution as the DHSs were generated by using the shuffle command in BEDTools. When multiple TE features were found for a single DHS, the longer TE feature was counted.

Data, Materials, and Software Availability. DNase-seq, ChIP-seq and RNA-seq for all cotton species used in this study have been deposited in the European Nucleotide Archive (ENA), https://www.ebi.ac.uk/ena (BioProject accession number PRJEB47222) (106). All other study data are included in the main text and supporting information.

ACKNOWLEDGMENTS. We thank Dr. Jonathan F. Wendel from Iowa State University for comments on the manuscript. We also thank Pengxi Wang for the technical support, Dr. Tianzhen Zhang from Zhejiang University for providing the wGh plant, and Xinlian Shen from Jiangsu Academy of Agricultural Sciences for providing the Gd plant. This work was supported by the National Natural Science Foundation of China (32070544), the National Key R&D Program of China (2021YFE0101200), and the Startup Foundation from Nantong University (03083074 and 135421609105). This work was also partially supported by Cotton Incorporated Grants 21-855 and 21-844; NSF Award 1658709; US Department of Agriculture–National Institute of Food and Agriculture Grants 2019-67029-35289 and 2021-67013-34738; and the State of Texas Governor's University Research Initiative/Texas Tech University Grant 05-2018 (to L.H.-E.).

Author affiliations: [a]School of Life Sciences, Nantong University, Nantong 226019, People's Republic of China; [b]Institute of Genomics for Crop Abiotic Stress Tolerance, Department of Plant and Soil Science, Texas Tech University, Lubbock, TX 79430; [c]College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, People's Republic of China; [d]Department of Biology, East Carolina University, Greenville, NC 27858; [e]Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, People's Republic of China; [f]National Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing 210095, People's Republic of China; and [g]Unidad de Genomica Avanzada/Laboratorio Nacional de Genómica para la Biodiversidad, Centro de Investigación y de Estudios Avanzados, Irapuato, México36821

Author contributions: D.L.-A., L.H.-E., B.Z., and K.W. designed research; J.H., D.L.-A., G.Y., Y.W., B.W., S.B.W., X.Z., H.F., A.C.B.-R., X.P., Y.J., J.C., H.Z., and B.Z. performed research; B.Z. and K.W. contributed new reagents/analytic tools; J.H., D.L.-A., G.Y., Y.W., B.W., S.B.W., X.Z., H.F., A.C.B.-R., X.P., Y.J., J.C., H.Z., B.-L.Z., L.H.-E., B.Z., and K.W. analyzed data; and J.H., G.Y., B.-L.Z., L.H.-E., B.Z., and K.W. wrote the paper.

Reviewers: M.L., Masarykova Univerzita Stredoevropsky Technologicky Institut; C.S., Clemson University; and Y.Z., Fudan University.

1. Y. Van de Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017).
2. J. F. Wendel, S. A. Jackson, B. C. Meyers, R. A. Wing, Evolution of plant genome architecture. *Genome Biol.* **17**, 37 (2016).
3. Z. J. Chen, Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* **58**, 377–406 (2007).
4. G. L. Stebbins Jr., Types of polyploids; their classification and significance. *Adv. Genet.* **1**, 403–429 (1947).
5. A. Madlung, J. F. Wendel, Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenet. Genome Res.* **140**, 270–285 (2013).
6. P. J. Wittkopp, B. K. Haerum, A. G. Clark, Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
7. Z. Lv *et al.*, Conservation and trans-regulation of histone modification in the A and B subgenomes of polyploid wheat during domestication and ploidy transition. *BMC Biol.* **19**, 42 (2021).
8. X. Jiang, Q. Song, W. Ye, Z. J. Chen, Concerted genomic and epigenomic changes accompany stabilization of *Arabidopsis* allopolyploids. *Nat. Ecol. Evol.* **5**, 1382–1393 (2021).
9. Q. Song, T. Zhang, D. M. Stelly, Z. J. Chen, Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. *Genome Biol.* **18**, 99 (2017).
10. A. Salmon, L. Flagel, B. Ying, J. A. Udall, J. F. Wendel, Homoeologous nonreciprocal recombination in polyploid cotton. *New Phytol.* **186**, 123–134 (2010).
11. M. Garcia-Lozano *et al.*, Altered chromatin conformation and transcriptional regulation in watermelon following genome doubling. *Plant J.* **106**, 588–600 (2021).
12. W. Zhu *et al.*, Altered chromatin compaction and histone methylation drive non-additive gene expression in an interspecific *Arabidopsis* hybrid. *Genome Biol.* **18**, 157 (2017).
13. T. Wicker *et al.*; International Wheat Genome Sequencing Consortium, Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biol.* **19**, 103 (2018).
14. H. Yan *et al.*, siRNAs regulate DNA methylation and interfere with gene and lncRNA expression in the heterozygous polyploid switchgrass. *Biotechnol. Biofuels* **11**, 208 (2018).
15. S. Henikoff, Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat. Rev. Genet.* **9**, 15–26 (2008).

16. J. Jiang, The 'dark matter' in the plant genomes: Non-coding and unannotated DNA sequences associated with open chromatin. *Curr. Opin. Plant Biol.* **24**, 17–23 (2015).
17. E. Birney *et al.*; ENCODE Project Consortium; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
18. S. Roy *et al.*; modENCODE Consortium, Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010).
19. S. J. Burgess *et al.*, Genome-wide transcription factor binding in leaves from C_3 and C_4 grasses. *Plant Cell* **31**, 2297–2314 (2019).
20. J. Han *et al.*, Genome-wide characterization of DNase I-hypersensitive sites and cold response regulatory landscapes in grasses. *Plant Cell* **32**, 2457–2473 (2020).
21. Z. Lu *et al.*, The prevalence, evolution and chromatin signatures of plant regulatory elements. *Nat. Plants* **5**, 1250–1259 (2019).
22. A. C. Barragán-Rosillo *et al.*, Genome accessibility dynamics in response to phosphate limitation is controlled by the PHR1 family of transcription factors in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2107558118 (2021).
23. W. Zhang *et al.*, High-resolution mapping of open chromatin in the rice genome. *Genome Res.* **22**, 151–162 (2012).
24. J. Sun *et al.*, Global quantitative mapping of enhancers in rice by STARR-seq. *Genomics Proteomics Bioinformatics* **17**, 140–153 (2019).
25. R. Oka *et al.*, Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* **18**, 137 (2017).
26. P. A. Fryxell, A revised taxonomic interpretation of *Gossypium* L. (Malvaceae). *Rheedea* **2**, 108–165 (1992).
27. D. Yuan *et al.*, Parallel and intertwining threads of domestication in allopolyploid cotton. *Adv. Sci. (Weinh.)* **8**, 2003634 (2021).
28. Z. J. Chen *et al.*, Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat. Genet.* **52**, 525–533 (2020).
29. J. F. Wendel, R. C. Cronn, Polyploidy and the evolutionary history of cotton. *Adv. Agron.* **78**, 139–186 (2003).
30. L. Fang *et al.*, Genomic analyses in cotton identify signatures of selection and loci associated with fiber quality and yield traits. *Nat. Genet.* **49**, 1089–1098 (2017).
31. T. Zhang *et al.*, Sequencing of allotetraploid cotton (*Gossypium hirsutum* L. acc. TM-1) provides a resource for fiber improvement. *Nat. Biotechnol.* **33**, 531–537 (2015).
32. M. J. Yoo, E. Szadkowski, J. F. Wendel, Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171–180 (2013).
33. M. Wang *et al.*, Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat. Plants* **4**, 90–97 (2018).
34. T. Zhao *et al.*, LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. *Genome Biol.* **19**, 195 (2018).
35. Y. Hu *et al.*, *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat. Genet.* **51**, 739–748 (2019).
36. X. Du *et al.*, Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nat. Genet.* **50**, 796–802 (2018).
37. J. A. Udall *et al.*, *De novo* genome sequence assemblies of *Gossypium raimondii* and *Gossypium turneri*. *G3 (Bethesda)* **9**, 3079–3085 (2019).
38. A. H. Paterson *et al.*, Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**, 423–427 (2012).
39. S. L. Klemm, Z. Shipony, W. J. Greenleaf, Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
40. P. J. Sabo *et al.*, Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat. Methods* **3**, 511–518 (2006).
41. A. P. Boyle *et al.*, High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
42. K. L. Bubb, R. B. Deal, Considerations in the analysis of plant chromatin accessibility data. *Curr. Opin. Plant Biol.* **54**, 69–78 (2020).
43. Y. Liu *et al.*, Histone H3K27 dimethylation landscapes contribute to genome stability and genetic recombination during wheat polyploidization. *Plant J.* **105**, 678–690 (2021).
44. Y. Guo, G. Qin, H. Gu, L. J. Qu, Dof5.6/HCA2, a Dof transcription factor gene, regulates interfascicular cambium formation and vascular tissue development in *Arabidopsis*. *Plant Cell* **21**, 3518–3534 (2009).
45. M. Zhuo, Y. Sakuraba, S. Yanagisawa, A jasmonate-activated MYC2-Dof2.1-MYC2 transcriptional loop promotes leaf senescence in *Arabidopsis*. *Plant Cell* **32**, 242–262 (2020).
46. V. Ruta *et al.*, The DOF transcription factors in seed and seedling development. *Plants* **9**, 218 (2020).
47. X. Wang *et al.*, Targeting of polycomb repressive complex 2 to RNA by short repeats of consecutive guanines. *Mol. Cell* **65**, 1056–1067.e5 (2017).
48. M. L. Theune, U. Bloss, L. H. Brand, F. Ladwig, D. Wanke, Phylogenetic analyses and GAGA-motif binding studies of BBR/BPC proteins lend to clues in GAGA-motif recognition and a regulatory role in brassinosteroid signaling. *Front. Plant Sci.* **10**, 466 (2019).
49. Y. Guo, S. Zhao, G. G. Wang, Polycomb gene silencing mechanisms: PRC2 chromatin targeting, H3K27me3 'readout', and phase separation-based compaction. *Trends Genet.* **37**, 547–565 (2021).
50. P. Li, Y. J. Lu, H. Chen, B. Day, The lifecycle of the plant immune system. *Crit. Rev. Plant Sci.* **39**, 72–100 (2020).
51. M. Hesami *et al.*, Advances and perspectives in tissue culture and genetic engineering of cannabis. *Int. J. Mol. Sci.* **22**, 5671 (2021).
52. K. M. Furuta *et al.*, Plant development. *Arabidopsis* NAC45/86 direct sieve element morphogenesis culminating in enucleation. *Science* **345**, 933–937 (2014).
53. S. Klees *et al.*, In silico identification of the complex interplay between regulatory SNPs, transcription factors, and their related genes in *Brassica napus* L. using multi-omics data. *Int. J. Mol. Sci.* **22**, 789 (2021).
54. Z. Li, S. H. Howell, Heat stress responses and thermotolerance in maize. *Int. J. Mol. Sci.* **22**, 948 (2021).
55. S. Quiroz *et al.*, Beyond the genetic pathways, flowering regulation complexity in *Arabidopsis thaliana*. *Int. J. Mol. Sci.* **22**, 5716 (2021).
56. R. Andersson, A. Sandelin, Determinants of enhancer and promoter activities of regulatory elements. *Nat. Rev. Genet.* **21**, 71–87 (2020).
57. Y. Lu, D.-X. Zhou, Y. Zhao, Understanding epigenomics based on the rice model. *Theor. Appl. Genet.* **133**, 1345–1363 (2020).
58. Q. Zhang *et al.*, Asymmetric epigenome maps of subgenomes reveal imbalanced transcription and distinct evolutionary trends in *Brassica napus*. *Mol. Plant* **14**, 604–619 (2021).
59. W. A. Ricci *et al.*, Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **5**, 1237–1249 (2019).
60. H. Zhao *et al.*, Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* **176**, 2789–2803 (2018).
61. E. Rodgers-Melnick, D. L. Vera, H. W. Bass, E. S. Buckler, Open chromatin reveals the functional maize genome. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3177–E3184 (2016).
62. M. Wang *et al.*, An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. *Plant Cell* **33**, 865–881 (2021).
63. R. Rebollo, M. T. Romanish, D. L. Mager, Transposable elements: An abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* **46**, 21–42 (2012).
64. Y. N. Ma *et al.*, Jasmonate promotes artemisinin biosynthesis by activating the TCP14-ORA complex in *Artemisia annua*. *Sci. Adv.* **4**, eaas9357 (2018).
65. Z. Xie, T. M. Nolan, H. Jiang, Y. Yin, AP2/ERF transcription factor regulatory networks in hormone and abiotic stress responses in *Arabidopsis*. *Front. Plant Sci.* **10**, 228 (2019).
66. J. Q. Yu *et al.*, The apple bHLH transcription factor MdbHLH3 functions in determining the fruit carbohydrates and malate. *Plant Biotechnol. J.* **19**, 285–299 (2021).
67. X. Guo *et al.*, Comparative transcriptome analysis of the floral transition in *Rosa chinensis* 'Old Blush' and *R. odorata* var. gigantea. *Sci. Rep.* **7**, 6068 (2017).
68. S. Ito *et al.*, FLOWERING BHLH transcriptional activators control expression of the photoperiodic flowering regulator CONSTANS in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 3582–3587 (2012).
69. L. Y. Wu *et al.*, Dynamic chromatin state profiling reveals regulatory roles of auxin and cytokinin in shoot regeneration. *Dev. Cell* **57**, 526–542.e7 (2022).
70. F.-X. Wang *et al.*, Chromatin accessibility dynamics and a hierarchical transcriptional regulatory network structure for plant somatic embryogenesis. *Dev. Cell* **54**, 742–757.e8 (2020).
71. H. S. Jang *et al.*, Epigenetic dynamics shaping melanophore and iridophore cell fate in zebrafish. *Genome Biol.* **22**, 282 (2021).
72. M. M. Halstead, X. Ma, C. Zhou, R. M. Schultz, P. J. Ross, Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. *Nat. Commun.* **11**, 4654 (2020).
73. L. Fang, X. Guan, T. Zhang, Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.). *Crop J.* **5**, 159–165 (2017).
74. Q. Li *et al.*, Unbiased subgenome evolution following a recent whole-genome duplication in pear (*Pyrus bretschneideri* Rehd.). *Hortic. Res.* **6**, 34 (2019).
75. H. Wu, Q. Yu, J. H. Ran, X. Q. Wang, Unbiased subgenome evolution in allotetraploid species of ephedra and its implications for the evolution of large genomes in gymnosperms. *Genome Biol. Evol.* **13**, evaa236 (2021).
76. J. G. Wallace *et al.*, Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genet.* **10**, e1004845 (2014).
77. M. T. Maurano *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
78. B. P. H. Metzger, P. J. Wittkopp, J. D. Coolon, Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol. Evol.* **9**, 843–854 (2017).
79. B. Lemos, L. O. Araripe, P. Fontanillas, D. L. Hartl, Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14471–14476 (2008).
80. D. R. Denver *et al.*, The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**, 544–548 (2005).
81. M. R. Corces *et al.*, The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
82. A. M. Sullivan *et al.*, Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep.* **8**, 2015–2030 (2014).
83. A. P. Marand, Z. Chen, A. Gallavotti, R. J. Schmitz, A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184**, 3041–3055.e21 (2021).
84. R. M. Graze *et al.*, Allelic imbalance in *Drosophila* hybrid heads: Exons, isoforms, and evolution. *Mol. Biol. Evol.* **29**, 1521–1532 (2012).
85. J. E. Endrizzi, E. L. Turcotte, R. J. Kohel, Genetics, cytology, and evolution of *Gossypium*. *Adv. Genet.* **23**, 271–375 (1985).
86. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory activities of transposable elements: From conflicts to benefits. *Nat. Rev. Genet.* **18**, 71–86 (2017).
87. C. Feschotte, Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **9**, 397–405 (2008).
88. A. Studer, Q. Zhao, J. Ross-Ibarra, J. Doebley, Identification of a functional transposon insertion in the maize domestication gene tb1. *Nat. Genet.* **43**, 1160–1163 (2011).
89. C. Zhou *et al.*, A genome doubling event reshapes rice morphology and products by modulating chromatin signatures and gene expression profiling. *Rice (N. Y.)* **14**, 72 (2021).
90. H. Zhang *et al.*, The effects of *Arabidopsis* genome duplication on the chromatin organization and transcriptional regulation. *Nucleic Acids Res.* **47**, 7857–7869 (2019).
91. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
92. W. Yan *et al.*, Dynamic control of enhancer activity drives stage-specific gene expression during flower morphogenesis. *Nat. Commun.* **10**, 1705 (2019).
93. Y. Wang, K. Wang, Genome-wide identification of DNase I hypersensitive sites in plants. *Curr. Protoc.* **1**, e148 (2021).
94. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
95. H. Li *et al.*; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

96. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
97. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
98. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
99. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
100. C. Trapnell *et al.*, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
101. J. Han *et al.*, Rapid proliferation and nucleolar organizer targeting centromeric retrotransposons in cotton. *Plant J.* **88**, 992–1005 (2016).
102. F. Ramírez *et al.*, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
103. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
104. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
105. S. Ou *et al.*, Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
106. J. Han *et al.*, DHSs maps dynamic accompanying genome duplication in cotton, PRJEB47222. ENA BioProject. https://www.ebi.ac.uk/ena/browser/view/PRJEB47222. Deposited 27 August 2021.