

# Transposable elements drive the subgenomic divergence of homoeologous genes to shape wheat domestication and improvement

Jizeng Jia<sup>1,2,9,\*</sup>, Pingchuan Deng<sup>3,4,9</sup>, Tianbao Li<sup>1</sup>, Kai Wang<sup>5</sup>, Lifeng Gao<sup>2</sup>, Guangyao Zhao<sup>2</sup>, Dangqun Cui<sup>1</sup>, Zhongdong Dong<sup>1</sup>, Chengdao Li<sup>6</sup>, Kehui Zhan<sup>1</sup>, Wanquan Ji<sup>4</sup>, Zhengang Ru<sup>7,\*</sup>, Daowen Wang<sup>1,\*</sup> and Liang Wu<sup>3,8,\*</sup>

<sup>1</sup>College of Agronomy, Collaborative Innovation Center of Henan Grain Crops, Henan Agricultural University, Longzi Lake Campus, Zhengzhou, China

<sup>2</sup>The National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

<sup>3</sup>Zhejiang Provincial Key Laboratory of Crop Genetic Resources, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China

<sup>4</sup>State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>5</sup>Novogene Bioinformatics Institute, Beijing, China

<sup>6</sup>Western Crop Genetics Alliance, College of Science, Health, Engineering and Education, Murdoch University, Perth, WA, Australia

<sup>7</sup>School of Life Science and Technology, Henan Institute of Science and Technology, Xinxiang, Henan, China

<sup>8</sup>Zhongyuan Institute, Zhejiang University, Zhengzhou, China

<sup>9</sup>These authors contributed equally to this article.

\*Correspondence: Jizeng Jia ([jiajizeng@caas.cn](mailto:jiajizeng@caas.cn)), Zhengang Ru ([rzgh58@163.com](mailto:rzgh58@163.com)), Daowen Wang ([dwwang@henau.edu.cn](mailto:dwwang@henau.edu.cn)), Liang Wu ([liangwu@zju.edu.cn](mailto:liangwu@zju.edu.cn))

<https://doi.org/10.1016/j.xplc.2025.101341>

## ABSTRACT

Polyploidization is a fundamental evolutionary process in plants, including bread wheat. In the present study, we performed a comprehensive genome-wide analysis of dynamic homoeologous gene divergence in Aikang58 (AK58), a modern elite polyploid wheat cultivar with a recently released reference genome, and in other wheat genomes, including landraces, synthetic wheat, and several breeding lines. Over 40% of transposable element (TE) families exhibit biased distribution across the three wheat subgenomes. Approximately 95.0% (113 421) of genes are co-located with TEs, and these variable TEs significantly contribute to homoeologous divergence. We found that about 80% of triad homoeologs are divergent due to differences in expression or sub-functionalization. In addition, subgenome divergence potentially promote polyploid wheat domestication and improvement by increasing favorable homoeoallele combinations. Our findings suggest that homoeolog divergence contributes to the adaptation, domestication, and improvement of hexaploid wheat. The contribution of subgenomic divergence to polyploid heterosis is also discussed. This study provides a valuable resource for the investigation of how TEs drive homoeologous divergence during wheat domestication and improvement.

**Key words:** subgenome gene divergence, polyploidy, transposable elements, wheat

Jia J., Deng P., Li T., Wang K., Gao L., Zhao G., Cui D., Dong Z., Li C., Zhan K., Ji W., Ru Z., Wang D., and Wu L. (2025). Transposable elements drive the subgenomic divergence of homoeologous genes to shape wheat domestication and improvement. *Plant Comm.* **6**, 101341.

## INTRODUCTION

Polyploidy, in which an organism or cell contains two or more complete sets of chromosomes, is an important and widespread evolutionary phenomenon in the plant kingdom. Approximately 50%–70% of angiosperms are polyploids (Coghlan et al., 2005). Notably, many important staple, fiber, and oil crops—such as bread wheat (*Triticum aestivum*,  $2n = 6 \times = 42$ , AABBDD), cotton (*Gossypium hirsutum*,  $2n = 4 \times = 52$ , AADD), oilseed rape (*Brassica napus*,  $2n = 4 \times = 38$ , AACCC), and peanut (*Arachis hypogaea*,

$2n = 4 \times = 40$ , AABB)—are allopolyploids that carry two or more related but non-identical subgenomes. Subgenome divergence and the dynamic behavior of polyploid genomes are essential aspects of polyploid evolution (Adams, 2007; Chen, 2007; Shi et al., 2012; Mayer et al., 2014; Cheng et al., 2018; International Wheat Genome Sequencing Consortium [IWGSC], 2018; Wang et al., 2021). Compared with their diploid progenitors, polyploids often exhibit polyploid vigor (Ni et al., 2009) and heterosis (Chen, 2010; 2013), which are key components of polyploid advantage (Wei et al., 2019). Accordingly, polyploids have been proposed

as model systems for the study of heterosis (Bansal et al., 2012; Washburn and Birchler, 2014).

Bread wheat is the most widely cultivated crop worldwide, serving as a staple food source for 35%–40% of the global population. A key factor in its success as a global food crop is its allohexaploid genome, which originated from two hybridization events involving three diploid species (*Triticum urartu* AA, *Aegilops speltoides*-like BB, and *Aegilops tauschii* DD). This genomic composition confers greater tolerance to a wide range of biotic and abiotic stresses than its diploid progenitors (Dubcovsky and Dvorak, 2007). Allopolyploidization and the successful development of artificial hexaploid wheat make bread wheat an excellent model for studying the effects of dynamic homoeolog divergence on phenotypic variation during polyploidization, domestication, and crop improvement (Adamski et al., 2020).

Transposable elements (TEs) are important components of medium to large genomes and are primary drivers of genome evolution and adaptation (Bourque et al., 2018; Baduel et al., 2019). Research suggests that polyploidization can induce bursts of TE transposition, likely due to incompatible transposition suppression mechanisms between the two donor genomes (Parisod et al., 2009; Vicient and Casacuberta, 2017; Ramachandran et al., 2020). Polyploidization may also increase the mutation rate and reduce the selective pressure on deleterious recessive mutations due to gene redundancy (Ronfort, 1999). Furthermore, the progressive accumulation of new TE insertions in polyploids can induce changes in genome structure and promote divergence between subgenomes and orthologous genes. Because TEs can carry “ready-to-function” *cis*-regulatory elements, TE insertions may influence gene expression by altering the regulation of adjacent genes (Ramachandran et al., 2020). However, TEs do not encode functional proteins maintained by natural selection, and it remains unclear whether biased insertion and accumulation lead to the loss of certain TE families. The wheat subgenomes are large (~5 Gb each) and contain a high proportion (>80%) of repetitive TEs (Zhao et al., 2017; Ling et al., 2018; Wicker et al., 2018), making polyploid wheat an excellent model system for studying the effects of TE activity on subgenome divergence.

A high-quality, chromosome-level reference genome was recently released for Aikang58 (AK58), an elite bread wheat variety widely cultivated in China (Jia et al., 2023). In this study, we comprehensively characterized subgenomic divergence in the AK58 genome by comparing it with the previously released genome of Chinese Spring (CS). We found that over 80% of triad homoeologs in AK58 are divergent, including 17.0% with predicted neo-functionalization and 76.4% with predicted subfunctionalization. We demonstrate that TEs are the primary drivers of homoeologous gene divergence in polyploid wheat and substantially contribute to polyploid advantage, as well as to wheat domestication and improvement.

## RESULTS

### TEs are co-located with genes

We identified 11 993 623 TE copies from 653 TE families in the AK58 genome. The total TE lengths in the A, B, and D subgenomes

(hereafter called subA, subB, and subD, respectively) were 4.2 Gb (86.7%), 4.5 Gb (85.5%), and 3.3 Gb (84.0%), respectively, similar to the content and composition of the CS genome (Figure 1A; Supplemental Table 1) but slightly different from those reported for the diploid progenitors: the *T. urartu* A genome (4.9 Gb, 81.4%) (Ling et al., 2018) and the *Ae. tauschii* D genome (4.3 Gb, 85.9%) (Zhao et al., 2017) (Supplemental Figure 1). To test the effect of TEs on gene divergence among the three wheat subgenomes, we investigated TE abundance in flanking regions and within gene bodies. A total of 94 214 genes (78.9%) contained TEs within 2 kb upstream of the transcription start site (TSS), 89 976 genes (75.3%) had TEs within 2 kb downstream of the transcription end site (TES), and 33 266 (27.8%) and 5930 genes (5.0%) contained TEs in intronic and exonic regions, respectively (Figure 1B; Supplemental Table 2). In total, 95.0% (113 421) of genes were co-located with TEs, with 111 716 (93.5%) having TEs in flanking regions and 39 196 (32.8%) in gene body regions.

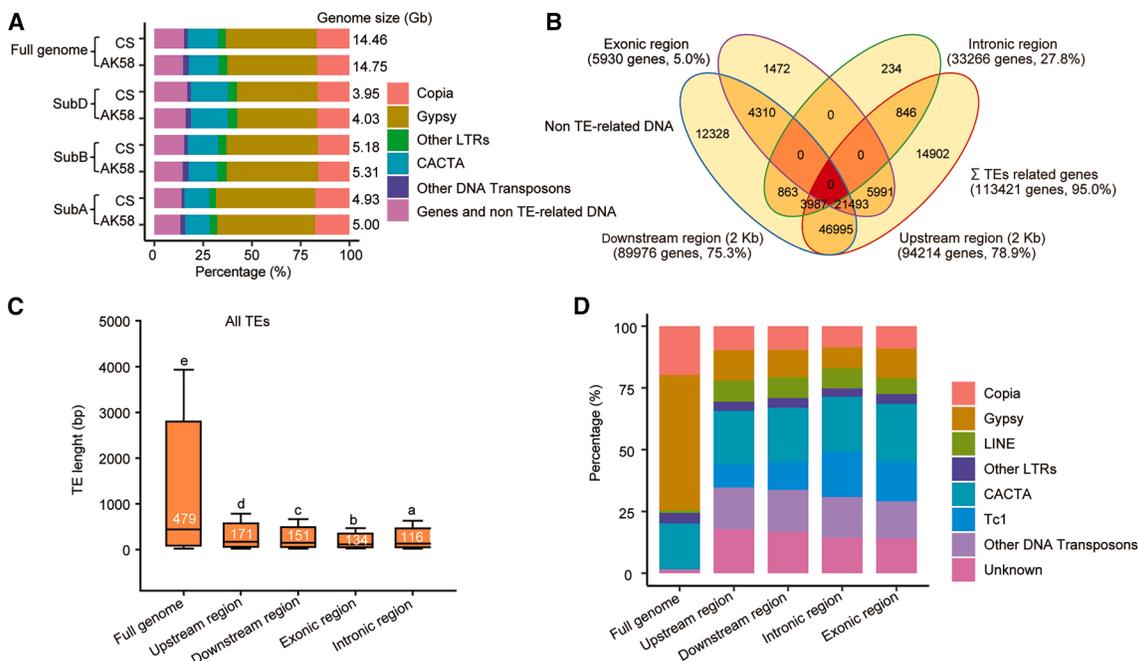
The lengths of TEs showed broad variations. Among the 2-kb upstream regions, 2-kb downstream regions, the median lengths were 171, 151, 116, and 134 base pairs (bp), respectively, accounting for 32.3%, 30.2%, 12.5%, and 21.1% of the base pairs in these regions. These median lengths were significantly shorter than the median TE length across the whole genome (Figure 1C). This difference was primarily due to a significant reduction in extended long terminal repeats (LTRs) in gene bodies and flanking regions, especially those from the LTR/Gypsy and LTR/Copia families (Supplemental Figure 2). By contrast, several miniature repeat element superfamilies were enriched within gene bodies or flanking regions, including DNA/Tc1, DNA/Mutator, and DNA/PIF. These results suggest that TEs are important components of wheat genes. Therefore, the variability of TEs and their co-location with genes and regulatory regions may drive subgenome gene divergence in terms of copy number, sequence, expression, and function.

We next compared TE distribution patterns in subA, subB, and subD. LTR, LINE, and Mutator elements exhibited significantly biased subgenomic distributions (Supplemental Table 3). In total, 645 TE families (42.1%) showed a bias in subgenome distribution, which potentially contributes to divergence among subgenomes and homoeologous genes. Further examination revealed distribution bias in a considerable proportion of TE families. Among the four types of TEs with even distributions across the three subgenomes, hAT and Tc1-Mariner elements showed no family-level distribution bias, whereas 9 of 41 PIF-Harbinger families and 13 of 23 CACTA families exhibited family-level bias.

### Effect of TEs on gene copy number divergence

#### Gene copy number divergence

We performed a phylogenomic analysis to identify gene homoeologs and paralogs in the AK58 and CS genomes. A total of 119 448 protein-coding genes (PCGs) were annotated in AK58 and classified into 4 categories: triads (54.4%), multiples (18.8%), dyads (8.7%), and singletons (18.1%) (Figure 2A; Supplemental Table 4). In contrast to triads, genes in the multiples category exhibited copy number variation (CNV), whereas dyads and singletons displayed presence/absence



**Figure 1. TE content and distribution in bread wheat.**

**(A)** Comparison of TE superfamily distributions between AK58 and CS across the three subgenomes.

**(B)** Proportions of genes co-located with TEs in gene body regions (exons and introns) and in 2-kb upstream and downstream flanking regions.

**(C)** Comparison of the length distributions of TEs co-located with genes. Statistical significance was assessed using ANOVA, and significant differences were found among the groups. Values represent the median TE size in each group.

**(D)** Proportions of the most abundant TE families co-located with genes.

variation (PAV) in subA, subB, or subD. Approximately 50% of the genes in the AK58 genome were affected by CNV or PAV across subgenomes. CNV- and PAV-affected genes showed significantly lower expression levels than triads and exhibited a trend toward differential expression between AK58 and CS (Figure 2B; Supplemental Table 5). Compared with triads, these genes also had fewer orthologs in other plant species (Figure 2C). Genes affected by CNV or PAV may contribute to genomic variability in higher plants.

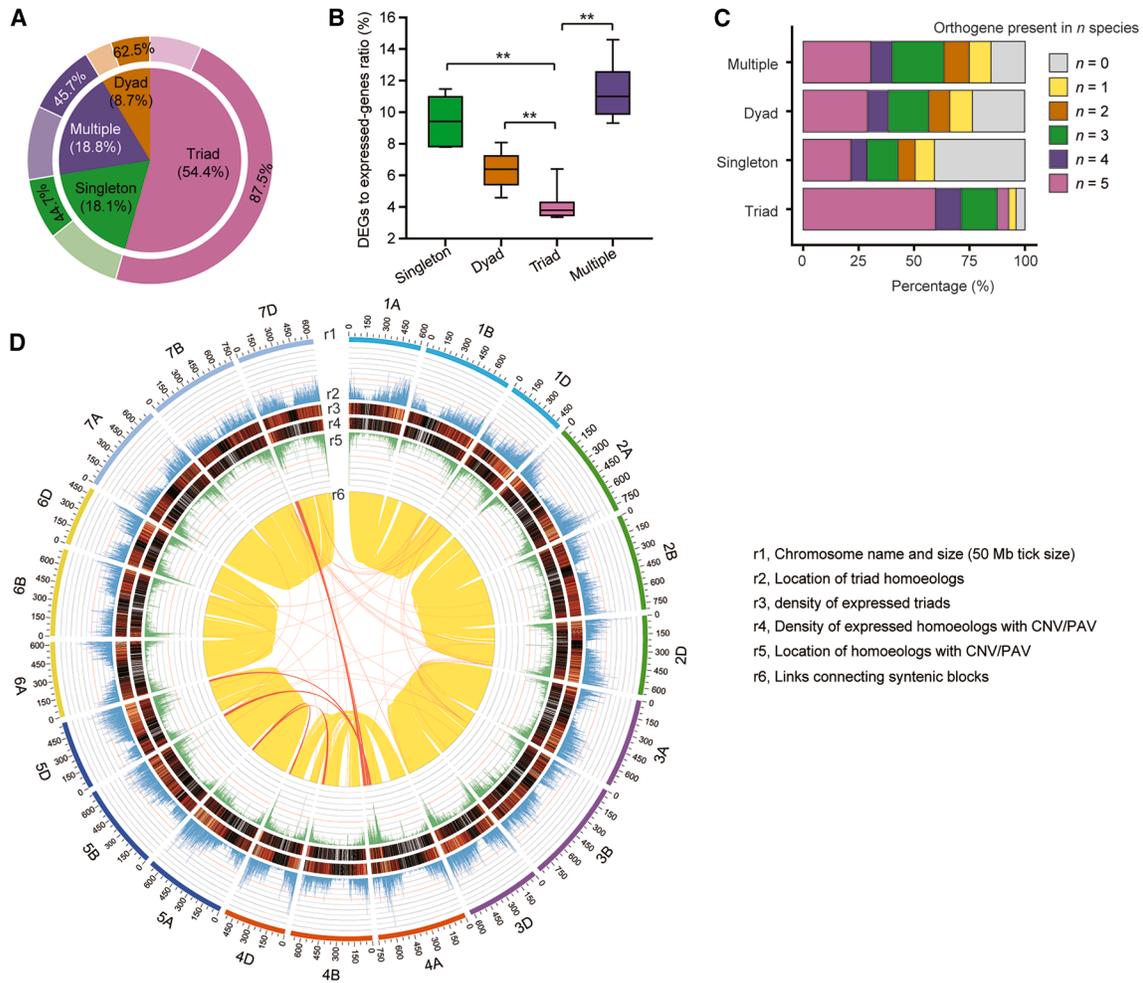
Genes in triads and those affected by CNV or PAV generally exhibited a mosaic distribution along chromosomes, with gene density decreasing from distal regions toward the centromere (Figure 2D). Using the collinearity between the CS and AK58 genomes, along with the previously defined genomic compartments of each CS chromosome (International Wheat Genome Sequencing Consortium [IWGSC], 2018), we divided the AK58 chromosomes into distal telomeric regions (R1 and R3), interstitial regions (R2a and R2b), and centromeric regions (C). We found a higher proportion of CNV- and PAV-affected genes in the telomeric regions (R1: 62.7%, R3: 53.2%) than in the interstitial (R2a: 40.2%, R2b: 32.2%) or centromeric regions (C: 40.0%) (Supplemental Figure 3). This pattern is likely explained by the fast-evolving nature of telomeric regions and the higher TE content in pericentric regions.

To investigate the origin of subgenome divergence, we compared PAV in AK58, CS, and ancestral wheat species. Among AK58 dyads, 73.5% of the missing homoeologs were also absent in both the ancestral species and CS

(Supplemental Figure 4A). Similar patterns were observed for dyads in CS (Supplemental Figure 4B) and for singletons in AK58 (Supplemental Figure 4C). These results suggest that most missing homoeologs were already absent in the ancestral species rather than being lost during allopolyploidization. Nevertheless, nearly 6% of missing homoeologs in both CS and AK58 were present in ancestral wheat species (Supplemental Figure 5), which indicates that some homoeolog losses may have occurred during allopolyploidization.

#### Effect of TEs on gene copy number variation (CNV)

In AK58, 95% of genes were co-located with TEs in their gene body or flanking regions ( $\pm 2$  kb). Genes co-located with TEs were more frequently found in triads and conserved in synteny than genes without TEs (Supplemental Figure 6A–6C). Transcriptome data showed that genes without TEs tended to have lower expression levels than those with TEs, which suggests that the ability to escape from TEs may be related to low transcriptional activity rather than low functional importance (Supplemental Figure 6D–6F). However, TE frequencies across the four gene categories differed significantly from the overall frequencies of the main TE families (Figure 3A and Supplemental Figure 7; Supplemental Table 6). Compared to triads, singleton genes exhibited the highest densities of CACTA, Gypsy, and Copia, the three longest and most abundant TE families. In contrast, shorter and rarer TE families—PIF, Tc1, and LINE—were detected more frequently in the flanking regions of triads and less frequently in singleton genes. In addition, comparisons of TE frequencies near genes showed that CACTAs were significantly enriched in the flanking regions of non-syntenic compared to syntenic genes (Supplemental Figure 8). AK58 is a typical wheat variety that



**Figure 2. Overview of homoeologs with CNV, PAV, or neo-functionalization in wheat across the three subgenomes.**

**(A)** Pie chart showing the categories of homoeologous genes in the AK58 genome. In the outer ring, darker segments indicate the fraction of expressed genes in each category.

**(B)** Genes with CNV or PAV are important types of differentially expressed genes (DEGs) between AK58 and CS. The ratio of DEGs to expressed genes is shown for each category. Wilcoxon–Mann–Whitney test,  $**p < 0.01$ .

**(C)** Homoeologs with CNV or PAV in the three wheat subgenomes are associated with fewer orthologs across five other species: *Hordeum vulgare*, *Oryza sativa*, *Brachypodium distachyon*, *Arabidopsis thaliana*, and *Glycine max*.

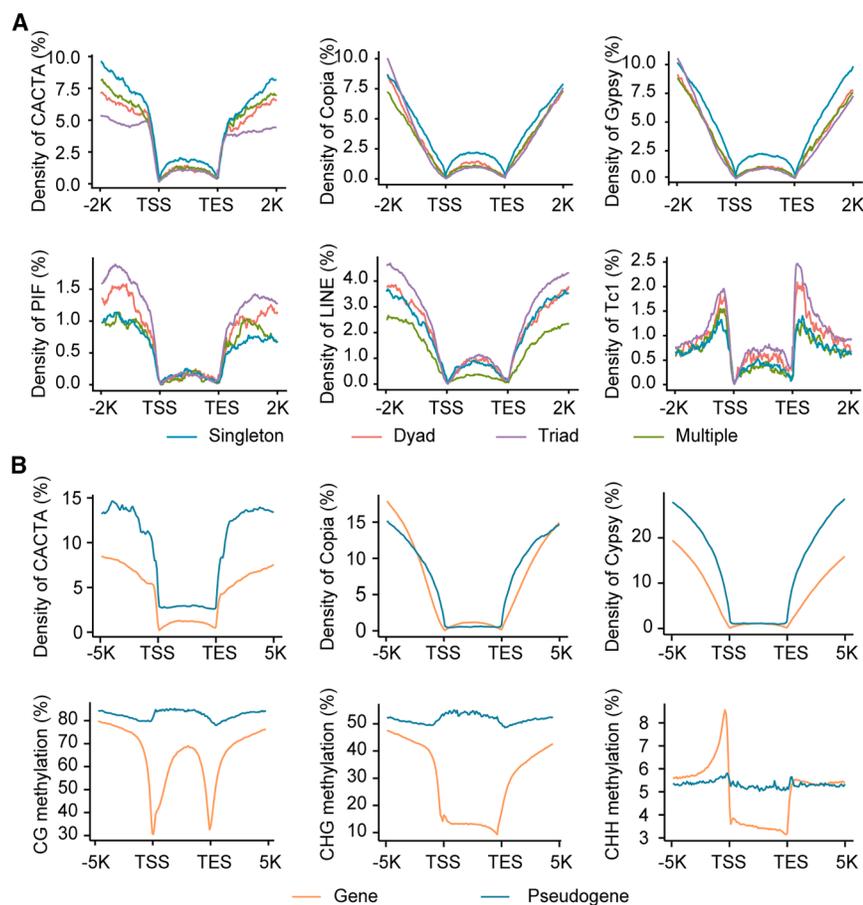
**(D)** Genomic structure, gene expression, and conserved syntenic landscape of homoeologs in wheat across the three subgenomes. Tracks from outer to inner ring: r1, chromosome names and sizes (50 Mb tick marks); r2, locations of triad homoeologs; r3, density of expressed triads (0–27 genes/Mb), with dark red indicating higher density; r4, density of expressed homoeologs with CNV or PAV (0–31 genes/Mb), with red indicating higher density; r5, locations of homoeologs with CNV or PAV; r6, links connecting syntenic blocks in AK58. Yellow links represent syntenic blocks across homoeologous group chromosomes, whereas red links indicate recombination.

carries the 1RS.1BL translocation; we found that genes on chromosome 1RS were more prone to CNVs than those on other chromosomes, which may be related to the higher abundance of PIF and Gypsy TEs in or near these genes (Supplemental Figure 9). Overall, these results suggest that TEs may significantly influence CNV.

TE-mediated gene capture contributes to CNV by affecting the duplication and movement of genes. CACTA-mediated gene capture has been reported in soybean (Zabala and Vodkin, 2005), sorghum (Paterson et al., 2009), *Ipomoea tricolor* (Takahashi et al., 1999), and wheat (Daron et al., 2014). We identified 2819 high-confidence (HC) genes linked to CACTA-mediated capture events in the AK58 genome (Supplemental

Table 7). The frequency of CACTA-mediated capture events was generally higher in subB and in the distal regions of chromosomes (Supplemental Figure 10). These capture events were distributed across 48 CACTA subfamilies (Supplemental Figure 11A) and were enriched in the multiple and singleton gene categories (Supplemental Figure 11B). There were 16 CACTA subfamilies with at least 50 captured genes, collectively accounting for 2518 (85.3%) of the captured genes. The most abundant subfamily was DTC\_famc5, which captured nearly 500 genes, accounting for 16.7% of all captured genes.

We analyzed gene duplication patterns in the AK58 genome using DupGen-finder (Qiao et al., 2019), which classifies duplicated genes into five categories: whole-genome duplication, tandem



duplication (TD), proximal duplication, transposed duplication, and dispersed duplication. A total of 583 CACTA-captured genes were classified as duplications, including 74 tandemly duplicated genes (Supplemental Figure 11C). Further analysis revealed that three CACTA subfamilies (DTC\_famc5, DTC\_famc6.1, and DTC\_famc1.2) were preferentially associated with gene duplications (Supplemental Figure 11D), whereas another five CACTA subfamilies (DTC\_famc2.2, DTC\_famc4.1, DTC\_famc4.3, DTC\_famc20, and DTC\_famc1.4) were associated with tandem gene duplication (Supplemental Figure 11E). Using currently available AK58 expression datasets, we found that 45.8% (1205) of the CACTA-captured genes were transcribed (Supplemental Figure 12A and 12B). Furthermore, the expressed CACTA-captured genes tended to be constitutively expressed (Supplemental Figure 12C) and were significantly enriched in biological processes related to lipid metabolism, organic acid metabolism, and isoprenoid metabolism (Supplemental Figure 12D).

A recent burst of gene duplications has been reported in Triticeae (Wang et al., 2022a). We found that CACTA elements were enriched in Triticeae relative to other species, which suggests that CACTA capture may have contributed to this recent burst. Among the 583 CACTA-captured duplicated genes, we observed enrichment in various families, including transcription factors, transcriptional regulators, and enzymes (Supplemental Figure 13A). A notable example is the *NAC* (*NAM/ATAF/CUC*) family, which is one of the largest groups of plant-specific

**Figure 3. Effect of TEs on gene copy number variation (CNV).**

**(A)** Density distributions of TEs in regions associated with subgenome homoeologs across various TE categories.

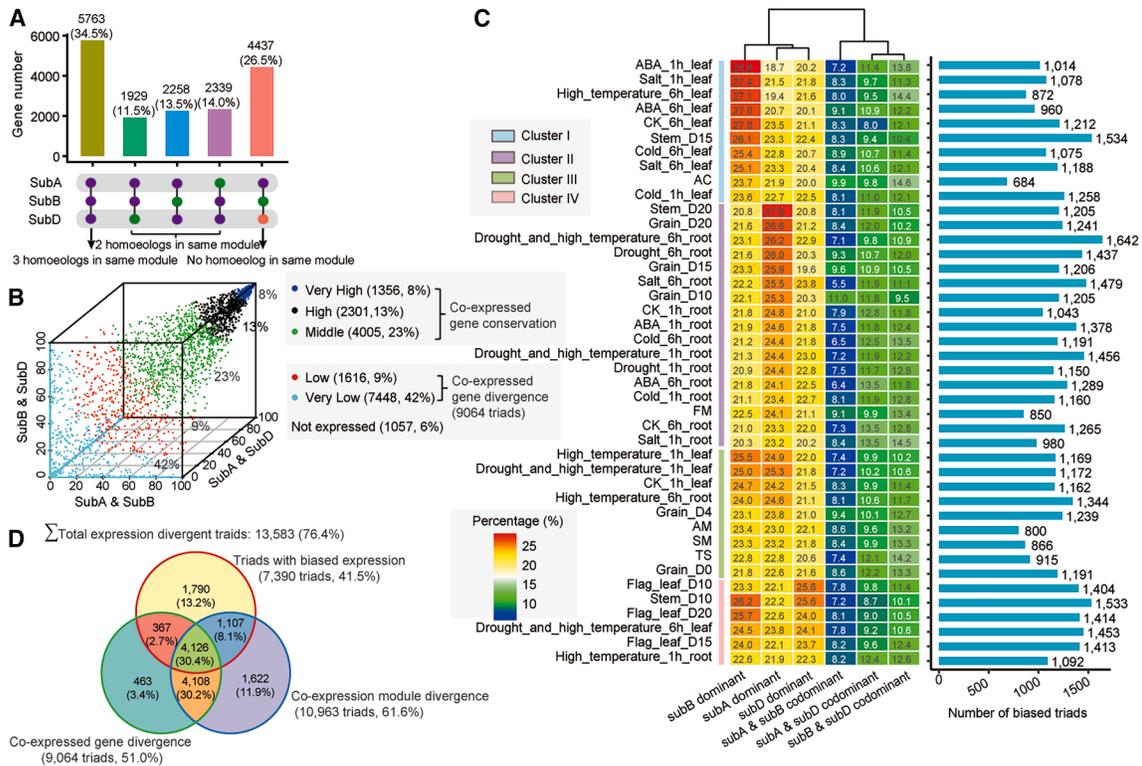
**(B)** Average TE densities and methylation levels surrounding genes and pseudogenes. TSS, transcription start site; TES, transcription end site.

transcription factors involved in a wide range of biological functions (Zhu et al., 2012). Around 100 *NAC* genes have been reported in diploid species, and the number was higher in AK58 (Supplemental Figure 13B). The AK58 genome contained 1.27 times as many *NAC* genes as CS and included 45 CACTA-captured *NAC* genes (Supplemental Figure 13C). Most of these CACTA-captured genes (38 genes, 84.4%) were classified into *NAC* clade I, which included 89 genes across the two genomes (1 in CS and 88 in AK58) and was characterized by a rapid rate of evolution (Supplemental Figure 13D). AGO (Argonaute) proteins, key players in all known small RNA-directed regulatory pathways (Vaucheret, 2008), were another typical example. The number of AGO genes in diploid wheat species was comparable to that found in other diploid plants but was less than one-third the number found in tetraploid wheat. Similar results were obtained for AK58, which contained 97 AGO genes, including 5 CACTA-captured AGO genes (Supplemental Figure 14). In addition, three of the five captured AGO genes showed evidence of expression, with preferential expression in flowers, suggesting an important role for CACTA-captured genes in wheat development.

Pseudogenization can also result in gene CNV among homoeologous genes. A total of 51 588 pseudogenes were annotated in the AK58 genome. TE density was much higher in pseudogenes than in PCGs (Figure 3B). Among the three most common TE superfamilies, CACTA and Gypsy elements were frequently inserted into gene bodies or flanking regions, which indicates that TEs may be the driving force in pseudogenization, resulting in gene CNV across the three subgenomes. We further compared TE and DNA methylation levels in pseudogenes across the three subgenomes. Pseudogenes generally exhibited higher DNA methylation levels than PCGs (Figure 3B), which supports a link between TE-induced pseudogenization and DNA methylation.

### Gene coding sequence divergence

Neo-functionalization refers to the evolutionary process by which duplicated genes functionally diverge at the annotation level (e.g., different Gene Ontology [GO] terms), whereas sub-functionalization describes duplicated genes that retain the same function but exhibit transcriptional divergence across



**Figure 4. Expression divergence of homeologous genes in AK58.**

**(A)** Module assignment of triads in the weighted gene coexpression network analysis (WGCNA) network.

**(B)** Distribution of triad conservation across five co-expression levels: very high ( $\geq 90\%$ ), high (70%–90%), medium (30%–70%), low (10%–30%), and very low ( $\leq 10\%$ ). The proportion of shared co-expressed genes was calculated for each gene in each triad and used to classify triads into these categories. Triads classified as low or very low were defined as exhibiting co-expression divergence.

**(C)** Heatmap of subgenome-dominant expression patterns across samples. The bar plot indicates the number of biased triads per sample.

**(D)** Diagram showing the number of homeologs in different categories that display divergence.

tissues or cell types (Roulin et al., 2013; Cheng et al., 2018). In AK58, 13.6% of triads (2414) were divergent in GO, PO (plant ontology), or TO (trait ontology) annotations, which suggests that these triads may be neo-functionalized (Supplemental Figure 15A and 15B). Among them, 666 triads showed divergence in predicted protein domains, and only 35 were under positive selection. Two of their rice orthologs, *OsDEG10* and *OsSodCc7*, have been shown to confer enhanced resistance to oxidative and salt stress (Park et al., 2009; Alfatih et al., 2020). These results suggest that most homeologous genes have similar functions and behave like alleles in a heterozygous diploid. We thus refer to them as “homoeoalleles.”

### Expression divergence of homeologs and the role of TEs

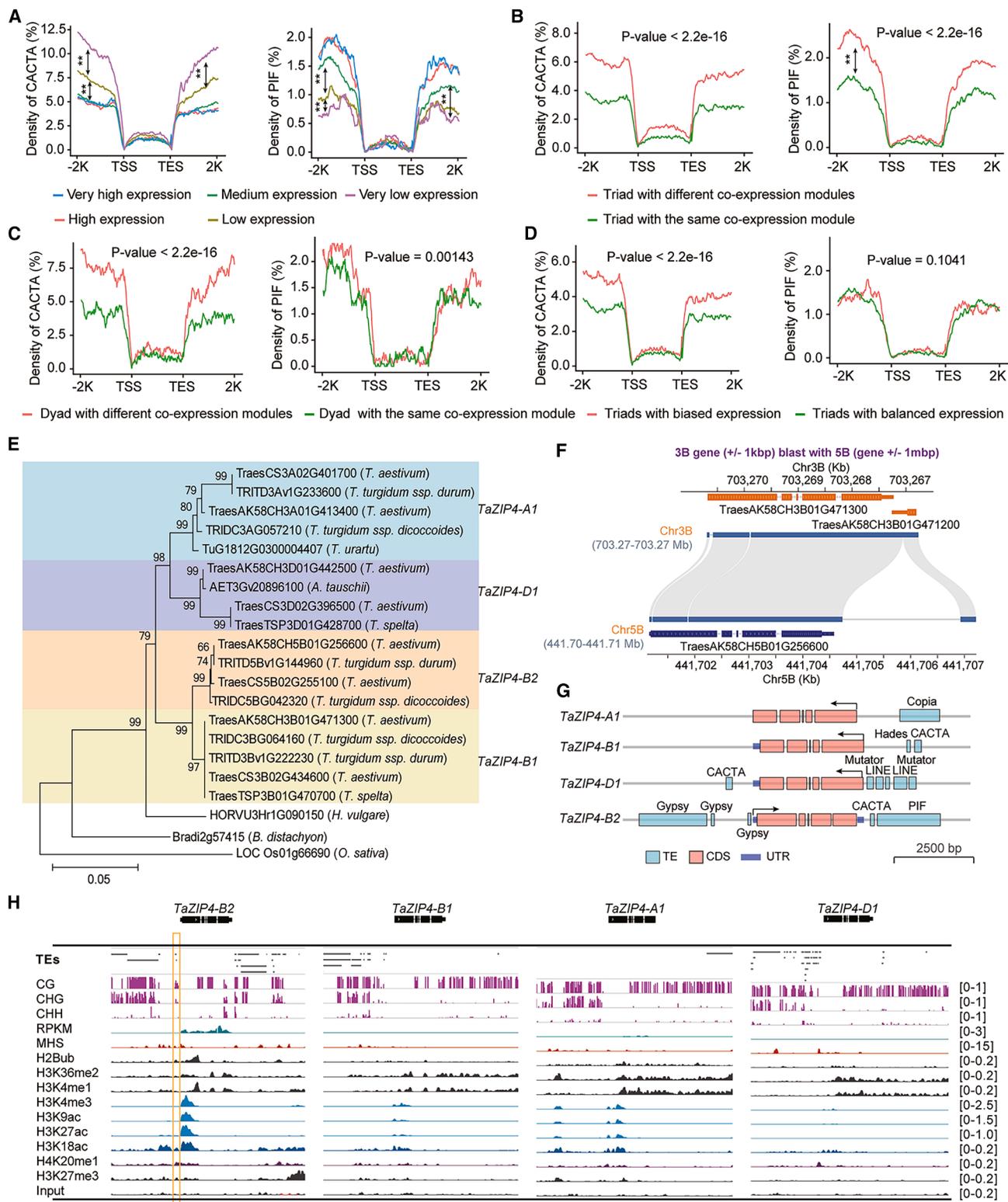
#### Variations in homoeologous gene expression

To explore the global gene expression landscape, we performed RNA sequencing (RNA-seq) on 42 diverse AK58 samples, including different tissues, developmental stages, and abiotic stress conditions (Supplemental Table 9). Of the 119 448 annotated PCGs, 82 704 (69.2%) were expressed, based on an abundance threshold of fragments per kilobase per million reads (FPKM)  $\geq 1$ . Because co-expressed genes are often functionally related (Pfeifer et al., 2014; Ramirez-Gonzalez et al., 2018), we constructed a co-expression network for the AK58 genes. Each of the 82 704 expressed genes was assigned to 1 of 84 co-

expression modules in the network (Supplemental Figure 16). To further assign module functions, we retrieved 2649 functionally characterized rice genes from the funRiceGenes database (<http://funricegenes.ncpgr.cn/>) to annotate the co-expression network.

We then analyzed divergence between homeologous genes using the gene co-expression network. Among the 17 783 syntenic triads identified in the AK58 genome, 94.1% (16 726) had at least 1 expressed homeologous transcript (Supplemental Figure 17). Among the expressed triads, 65.5% (10 963) exhibited divergence in their co-expression modules (termed “co-expression module divergence”); 4437 showed highly divergent expression (with no other homeologs in the same module), and 6526 showed an intermediate level of divergence (with 2 homeologs in the same module) (Figure 4A). In general, greater expression divergence was observed between homeologs from subA and subB, consistent with previous findings from CS transcriptome analysis (Pfeifer et al., 2014). Compared to homeologs in the same module, those in divergent modules tended to have narrower expression breadth and lower expression levels (Supplemental Figure 18A and 18B).

Given that homeologs can display divergent expression patterns even when grouped into the same co-expression module (Ramirez-Gonzalez et al., 2018), we identified the set of co-expressed genes for each homeolog and determined the



**Figure 5. Effect of TEs on the divergence of homoeologous gene expression.**

**(A)** Distribution of DNA/CACTA and DNA/PIF TEs in indicated regions of genes with very high ( $\geq 90\%$ ), high (70%–90%), medium (30%–70%), low (10%–30%), and very low ( $\leq 10\%$ ) expression levels.

**(B and C)** Density distributions of DNA/CACTA and DNA/PIF TEs in the regions surrounding triads **(B)** or dyads **(C)** assigned to different co-expression modules.

(legend continued on next page)

proportion of shared co-expressed genes for pairwise comparisons within each homoeologous group. We found that 51.0% (9064 of 17 783) of syntenic triads exhibited a proportion of less than 30% (defined as “co-expression gene divergence”) (Figure 4B), although 9.1% (829 of 9064) of these were grouped into the same co-expression module (Supplemental Figure 18C). In addition, we observed that homoeologs affected by rearrangement, PAV, or CNV exhibited a higher proportion of co-expression gene divergence than those in syntenic triads (Supplemental Figure 19), which supports the hypothesis that gene rearrangement and duplication can be followed by expression divergence, resulting in subsequent sub-functionalization (Chi et al., 2011).

Among the 16 726 triads with at least 1 homoeolog expressed in 1 of the 42 samples, we found that 7390 (44.2%) exhibited homoeolog-biased expression (Benjamini–Hochberg [BH]-corrected  $p < 0.05$  and fold change  $\geq 1.5$ ) in at least one condition. Across individual conditions, the proportion of biased triads ranged from 3.8% (anther connective stage) to 9.2% (seedling roots under drought and high temperature) (Figure 4C and Supplemental Figure 20). Furthermore, 75.8% of the biased triads (5600 of 7390) exhibited divergence in co-expression modules or co-expressed genes (Figure 4D). Overall, 76.4% of the AK58 triads showed expression divergence based on the expression data from the 42 samples, meaning that over three-quarters of the triads have varying degrees of expression divergence, which may reflect sub-functionalization and possibly lead to functional innovations. On chromosome 1RS, 484 genes were grouped into triads. We identified 222 triads (45.9%) that exhibited subgenome-biased expression, among which 131 had their dominantly expressed gene located on chromosome 1RS. These included genes with rice orthologs associated with disease resistance (*OsXa5* and *OsPti1a*), abiotic stress response (*OsPLC4*, *OsWRKY53*, and *OsDHAR*), and yield (*OsRSR1* and *OsECS*) (Supplemental Table 10). These genes may provide clues for further investigation of favorable homoeoalleles on 1RS to support wheat improvement.

We identified a total of 8059 single-copy orthologs in AK58, CS, and wheat ancestors. We then compared their expression patterns using transcriptomes from 16 newly synthesized hexaploid (SH) wheat lines and their donor parents. This dataset included samples from three tissues (root, leaf, and grain), one abiotic stress condition (cold stress), and two selfed generations (S1 and S5) (Supplemental Table 11). Pairwise comparisons showed that approximately 65.5% of biased triads in SH also exhibited biased expression in bulk samples from the SH donor parents (Bulk-SHDP) (Supplemental Figure 21). The remaining 34.5% displayed balanced expression patterns in the parental species. These findings suggest that polyploidy-biased gene

expression is not solely inherited from the original parents but is also generated during the process of polyploidization.

#### Effect of TEs on homoeologous gene expression divergence

To further investigate the influence of neighboring TEs on homoeologous gene expression, we classified coding genes into five categories based on their expression levels, ranked by percentile: very high ( $\geq 90\%$ ), high (70%–90%), medium (30%–70%), low (10%–30%), and very low ( $\leq 10\%$ ). We found that CACTA, Gypsy, and Copia, the top three TE superfamilies, were enriched in the upstream and downstream regions of genes with low expression but were underrepresented near highly expressed genes (Figure 5A and Supplemental Figure 22), which suggests that these TEs may suppress adjacent gene expression. This suppression may be associated with the positive correlation between TE abundance and levels of CG and CHG methylation near promoter regions, as previously reported in apple (Daccord et al., 2017) and soybean (Song et al., 2013).

We observed that certain TEs, including members of the Tc1 and PIF families, exhibited significant enrichment in highly and very highly expressed genes relative to other categories (Figure 5A and Supplemental Figure 22). This suggests that these TEs may carry ready-to-function regulatory elements, such as enhancers, that influence gene expression. Therefore, we searched for *cis*-elements in TEs (*TE-cis*) within 2-kb upstream regions and identified 93 104 genes that contained *TE-cis* elements. Among the 17 783 syntenic triads, 66.2% (11 782) were co-located with different *TE-cis* elements in at least 1 pair of homoeologous genes. Variation in *TE-cis* elements derived from different TE insertions could influence gene expression levels, generate novel promoters, and produce alternative TSSs.

We analyzed the TE landscape of wheat homoeologs with divergent expression patterns. Both triads and dyads with co-expression divergence had relatively high densities of CACTA, PIF, Mutator, Tc1, Copia, and Gypsy family TEs in their 2-kb upstream and downstream regions (Figure 5B and 5C and Supplemental Figures 23–25). In addition, higher densities of CACTA TEs were found in the promoters of triads with biased expression compared to those without biased expression (Figure 5D and Supplemental Figure 26).

ZIP4 (*ph1*) is a tetratricopeptide repeat protein that assembles protein complexes promoting homologous crossover during plant meiosis (Griffiths et al., 2006). This gene was originally located in homoeologous group 3 (3A, 3B, and 3D; 1:1:1; defined as *TaZIP4-A1*, *TaZIP4-B1*, and *TaZIP4-D1*, respectively) based on ortholog analysis. Another copy was mapped to chromosome 5B (0:1:0; defined as *TaZIP4-B2*). No *TaZIP4-B2* homoeologs were found in subA or subD. Phylogenetic analysis

(D) Density distribution of DNA/CACTA in regions surrounding triads with biased expression. The Wilcoxon–Mann–Whitney test was used to compare the distributions of expression levels between categories. \*\* $p < 0.01$ , \* $p < 0.05$ .

(E) Phylogenetic tree of *TaZIP4* genes.

(F) Reciprocal BLAST results for sequences  $\pm 1$  Mb around *TaZIP4-B2* and *TaZIP4-B1*.

(G) Gene structures and co-located TEs of *TaZIP4* genes.

(H) Effect of TEs on the pairing gene *TaZIP4*. This genome browser view displays TEs, DNA methylation (CG, CHG, and CHH), gene expression (FPKM-normalized), methylated histone sites, and histone modifications for *TaZIP4* genes on homoeologous group 5 (*TaZIP4-B2*) and group 3 (*TaZIP4-B1*, *TaZIP4-A1*, *TaZIP4-D1*) chromosomes in the AK58 genome. The orange rectangle indicates TEs in the *TaZIP4-B2* promoter region associated with DNA methylation, chromatin accessibility, and active histone marks.

indicated that *TaZIP4-B2* was closer to *TaZIP4-B1* than to *TaZIP4-A1* or *TaZIP4-D1* (Figure 5E). Moreover, a 200-bp fragment downstream of *TaZIP4-B2* was uniquely collinear with a gene adjacent to *TaZIP4-B1* (Figure 5F). Therefore, *TaZIP4-B2* likely originated from *TaZIP4-B1*, possibly after the divergence of emmer and spelt (Figure 5E). Most gene translocations are induced by TEs, either through TE-mediated gene capture and translocation or mispairing. Three LTR/Gypsy TEs and four other TEs (one DNA/CACTA, two DNA/PIF, and one unknown) were identified in the 2-kb upstream and downstream regions of *TaZIP4-B2*, respectively (Figure 5G and 5H), which suggests that its translocation might be related to these TEs. In AK58, *TaZIP4-B2* exhibited higher expression levels than *TaZIP4-A1*, *TaZIP4-B1*, and *TaZIP4-D1* across all tested tissues and conditions (Supplemental Figure 27A). This differential expression contributed to co-expression divergence among *TaZIP4* homoeologs (Supplemental Figure 27B) and may enhance homologous pairing. We also compared chromatin accessibility and epigenetic modifications among these genes. H3K18ac-induced transcriptional activity was observed at the TEs in the promoter region of *TaZIP4-B2* but not at the *TaZIP4* copies in homoeologous group 3 (Figure 5H). An LTR/Gypsy element co-located with open and active peaks at the promoter of *TaZIP4-B2*, whereas no TEs were found in the promoters of the other copies. The differential expression between *TaZIP4-B2* and the *TaZIP4* copies from homoeologous group 3 likely resulted from promoter alterations mediated by TE-induced translocation.

#### Effect of TEs on variation in DNA methylation of homoeologous genes

DNA methylation is a stably inherited epigenetic marker that alters chromatin density and DNA accessibility, affecting nearby transposon and gene activity as well as genome integrity. Previous research has indicated that most methylation occurs within TE regions (Li et al., 2019). In the present study, we generated whole-genome DNA methylation data for AK58 at single-base resolution. We found that 116 625 genes, accounting for 97.6% of all annotated genes, exhibited methylation in their promoters or gene bodies. To investigate how TEs and DNA methylation interact in different contexts to influence nearby gene activity, we characterized the genomic distribution of DNA methylation relative to TE type and density, as well as the relationship between methylation levels and the expression levels of nearby genes. We observed that DNA methylation in surrounding genes displayed different patterns in different contexts. The main methylation types in AK58 were CG and CHG, which were positively correlated with TE abundance in promoter regions (Figure 6A). In contrast, CHH methylation showed no clear relationship with TE density (Figure 6B).

To examine the effects of TEs and DNA methylation on gene divergence among subgenomes, we categorized the 116 625 methylation-related genes into 5 major homoeologous groups (Figure 6C) and examined TE density and DNA methylation for each group. Triads displayed higher levels of DNA methylation, especially non-CG methylation, than non-triplet genes (Figure 6D). The abundance of LTR TEs was slightly higher in the promoters of triads than in those of non-triplet genes (Figure 6E). We also calculated the methylation levels surrounding triads and compared them across subgenomes.

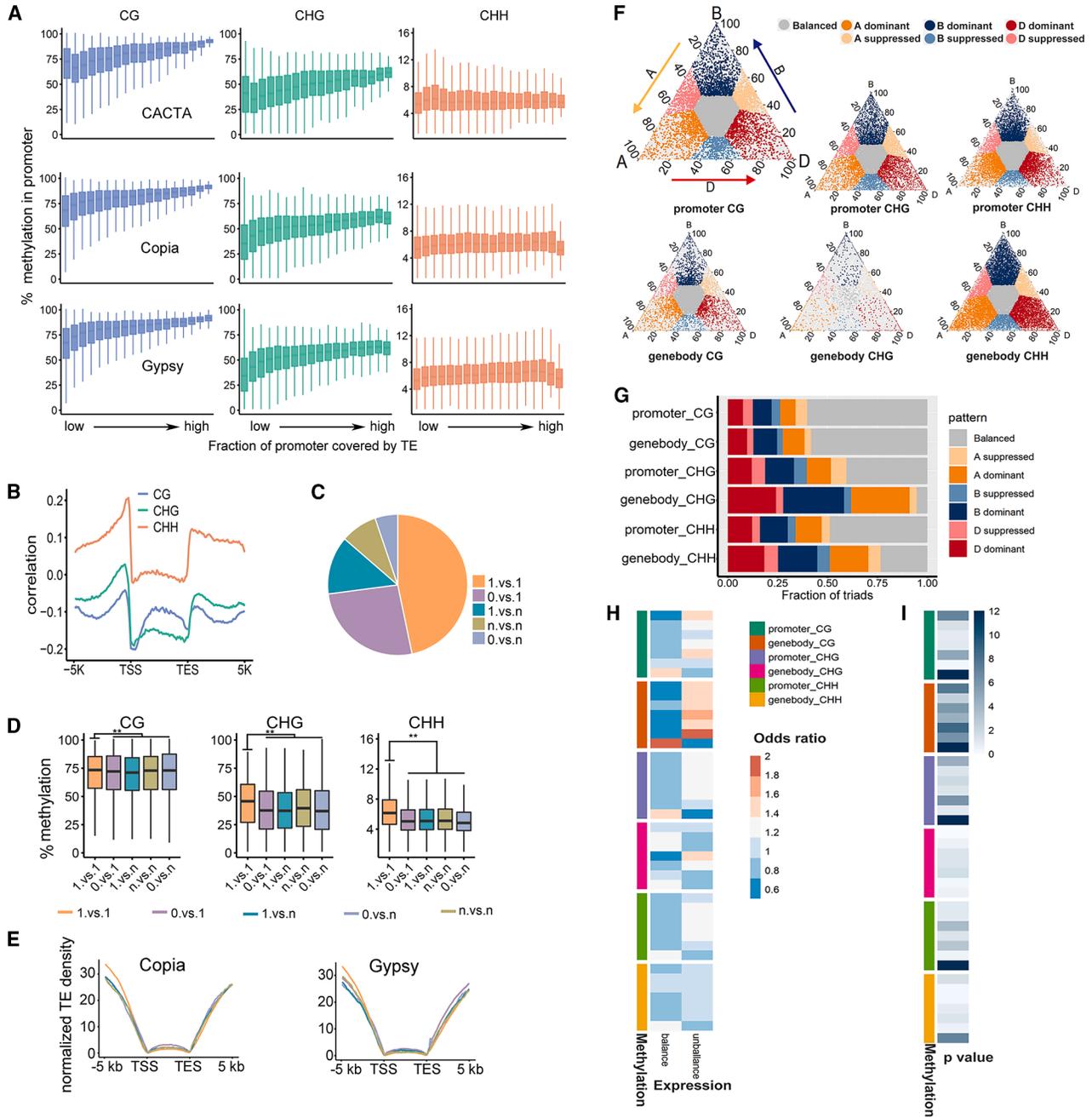
Triads were grouped into seven categories based on DNA methylation levels across three contexts (CG, CHG, and CHH): one group with similar methylation levels across the three homoeologs and six groups showing dominance or suppression in a single homoeolog, indicated by higher or lower methylation levels, respectively (Figure 6F). The fraction of triads in each category is shown in Figure 6G. More than 99% of triads exhibited divergence in DNA methylation, with only 69 triads showing no divergence in either the promoter or gene body regions (Supplemental Table 12). These results suggest that DNA methylation is significantly correlated with TEs and plays an important role in homoeologous gene divergence, possibly influencing expression variation (EV) among homoeologs (Figure 6H and 6I).

#### Homoeolog divergence contributes to polyploid advantage by increasing functional gene diversity and expression plasticity

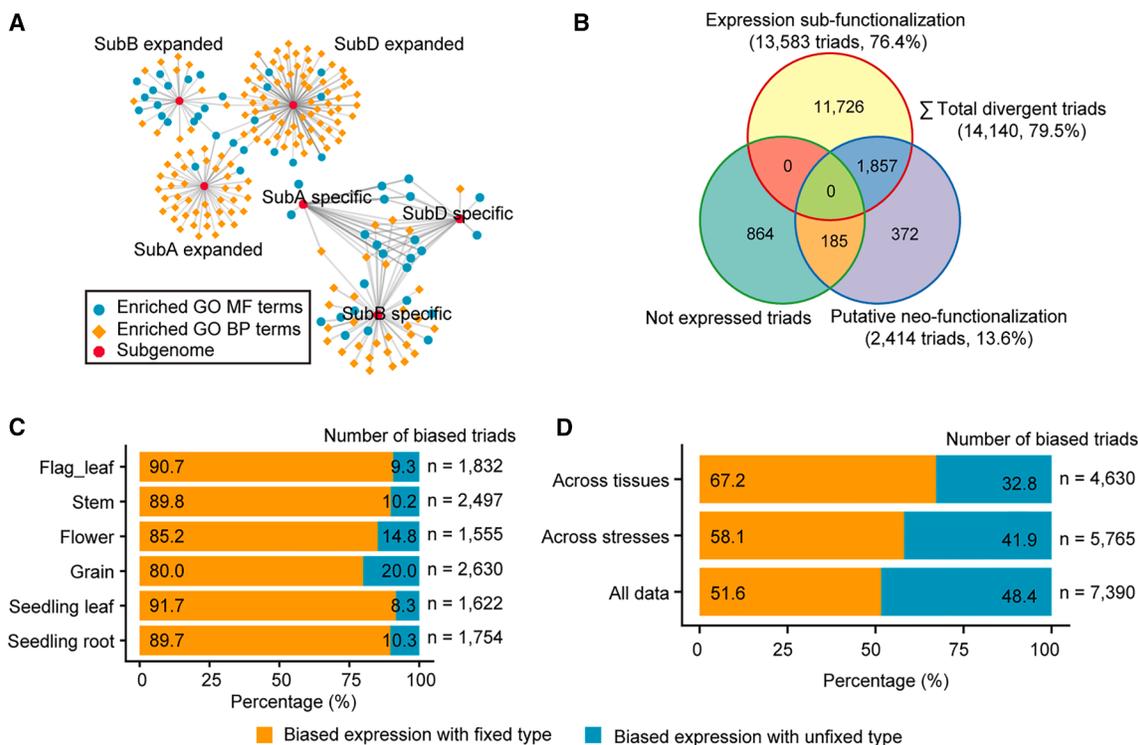
CNV and PAV likely contribute to crop phenotypic diversity (Springer et al., 2009). In the AK58 genome, they occurred in 53 475 genes across the three subgenomes, including 44.0% in subA, 45.0% in subB, and 47.6% in subD. GO analysis showed that overrepresented terms differed across subgenomes, regardless of whether duplicate genes were present in a specific subgenome ( $n:1:1$ ,  $1:n:1$ , or  $1:1:n$ , where  $n > 1$ ) (Figure 7A and Supplemental Figures 28 and 29). For instance, genes specific to subB were enriched in ATP metabolic process GO terms, whereas subD-specific genes were enriched in transferase activity terms.

In the AK58 genome, nearly 13.6% of triad homoeologs were putatively neo-functionalized; this refers to the evolutionary process by which duplicated genes diverge in function (Roulin et al., 2013; Cheng et al., 2018) (Figure 7B and Supplemental Table 8). More triads were sub-functionalized than neo-functionalized, exhibiting either divergent network connections or subgenome-biased expression (76.4% vs. 13.6%) (Figure 7B). These results suggest that functionally divergent genes are prevalent in polyploid wheat, which may have fueled wheat evolution, domestication, and improvement (Adams and Wendel, 2005; Van de Peer et al., 2009).

Allele-specific expression, or imbalanced expression of two parental alleles in a hybrid, has been considered a trigger for heterosis (Shao et al., 2019). Overall, we identified 7390 homoeologous triads (44.2%) that displayed biased expression (Supplemental Table 13). Two types of biased triads were identified across diverse transcriptome data: those with unfixed bias, where the type of bias varies between tissues or conditions, and those with fixed bias, where the type of bias is consistent across tissues and conditions. In transcriptome datasets from samples at different developmental stages, the proportions of triads showing fixed or unfixed bias averaged 87.9% and 12.1%, respectively (Figure 7C). In contrast, under stress conditions, 41.9% of the biased triads exhibited unfixed bias (Figure 7D). These biased triads were associated with functional terms related to stimulus responses and energy metabolism (Supplemental Figure 30A and 30B), and may contribute to the environmental adaptability of polyploid wheat. In addition, triads showing unfixed bias exhibited higher



**Figure 6. Relationship between subgenome-biased DNA methylation and gene expression.**  
**(A)** Distribution of DNA methylation levels in the CG, CHG, and CHH contexts within promoter regions (5 kb upstream of the TSS), grouped by TE coverage. Genes were categorized into 20 groups based on the fraction covered by CACTA, Copia, or Gypsy, the three most abundant TE types in wheat.  
**(B)** Correlation between 24-nt small RNA density and DNA methylation levels in the CG, CHG, and CHH contexts around genes.  
**(C)** Pie chart showing the proportions of genes in different homoeologous groups. The four non-triplet groups include: one copy in one subgenome and multiple paralogs in other subgenomes; n vs. n, multiple paralogs in different subgenomes; 1 vs. 0, only one copy present in one or two subgenome(s); and 1 vs. n, one copy in one or two subgenome(s) and multiple paralogs in other subgenome(s).  
**(D)** Distribution of promoter methylation levels for the homoeologous groups shown in **(A)**.  
**(E)** Distribution of TE density surrounding the homoeologous groups shown in **(A)**.  
**(F)** Ternary plot showing relative DNA methylation bias between triplet homoeologs. Each circle represents a triad. The Euclidean distance between the triad and the three vertices reflects the relative methylation level in the homoeologs. Triads were classified into seven methylation patterns based on Euclidean distance. A gene body or promoter with higher methylation in the A subgenome than the others was defined as A-dominant, whereas lower methylation in the A subgenome was defined as A-suppressed.  
**(G)** Proportion of triads with promoter or gene body regions exhibiting the seven methylation patterns shown in **(A)**. DNA methylation in the CG, CHG, and CHH contexts was calculated separately.  
**(H)** Effect of TEs on expression variation (EV) among homoeologs.



**Figure 7. Contribution of homoeologous gene divergence to polyploid advantage.**

**(A)** GO term network enrichment among subgenome-specific and expanded gene families.

**(B)** Diagram of triad homoeologs displaying putative sub-functionalization and neo-functionalization.

**(C)** Percentage of triads with biased expression patterns across different tissues.

**(D)** Percentage of triads with biased expression patterns across tissues and stress conditions. A triad is classified as showing fixed bias when all analyzed samples exhibit the same type of bias and unfixed bias when more than one type of bias is observed. Note that this classification is context-dependent and may vary if the tissues or conditions are changed.

co-expression divergence than triads with fixed bias or unbiased triads (Supplemental Figure 30C).

The dominantly expressed homoeologs were ranked by expression level across all transcriptome samples, which revealed that biased expression generally occurs in specific tissues and conditions. More than 81% of the 7390 biased triads that we identified showed homoeolog expression dominance in at least one of the samples with the 5 highest expression levels, and 42% showed biased expression in the sample with the highest overall expression (Supplemental Figure 31). These results indicate that biased expression typically occurs during the most active stages of gene expression, which suggests that under such conditions, the corresponding homoeolog may be well-adapted to the environment through long-term evolutionary selection.

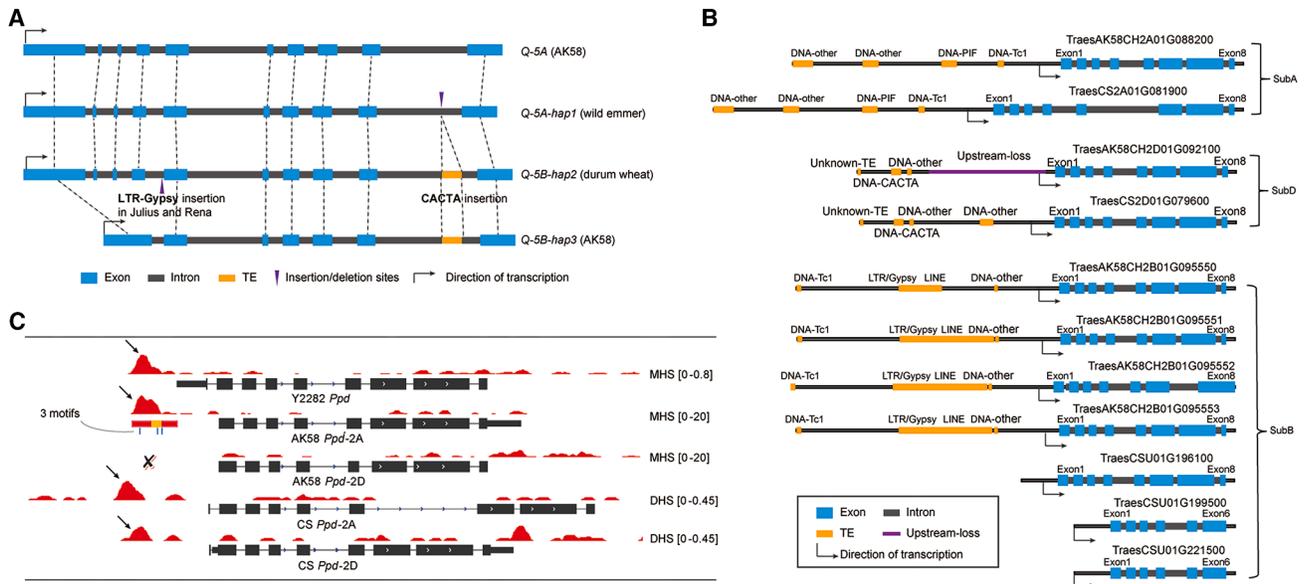
A typical example is the glutathione S-transferase (GST) protein family, a ubiquitous enzyme family that functions in cellular detoxification and environmental stress tolerance (Edwards et al., 2000; Wang et al., 2020). Among the 40 sets of homoeologous loci carrying GST genes in the AK58 genome (Supplemental Figure 32), 32 (80.0%) showed homoeolog-biased expression, with 25 (78.1%) of them varying across tissues and stress conditions (Supplemental Figure 33). This variation may be associated with the distinct substrate specificities and kinetic characteristics of different GST proteins, thereby facilitating the adaptation of

polyploid wheat to various biotic and abiotic challenges (Labrou et al., 2015).

### Homoeologous divergence contributes to wheat domestication and improvement

Many agronomically important genes have been cloned in wheat (Xiao et al., 2022). We collected 377 cloned triads and analyzed their homoeologous divergence (Supplemental Table 14). Of the 377 genes, 299 (79.3%) were divergent, with 293 showing divergence in expression and only 6 (1.6%) showing divergence with respect to GO terms. Therefore, most homoeologs have similar functions and exhibit sub-functionalization, which is analogous to the behavior of alleles in heterozygous diploids (i.e., they act as homoeoalleles). We further examined several well-known genes, including the domestication genes *Btr1* and *Q* and the improvement genes *Ppd1* and *Rht1*, to illustrate the functional divergence of homoeologous genes.

Rachis brittleness is a key trait in wheat domestication. Wild species in Triticeae have brittle rachis (BR), whereas domesticated wheat has a non-brittle rachis. BR is conferred by *Btr1* and *Btr2*, located on the short arms of the homoeologous group 3 chromosomes. Recessivity and loss of function of these two genes result in a spike non-shattering phenotype (Pourkheirandish et al., 2018). We found that 11 *Btr1* and 14 *Btr2* genes were located on



**Figure 8. Dynamics of TE-associated genes from diploid ancestors to hexaploids.**

**(A)** Overview of the intron/exon structure of *Q-5A* and *Q-5B* in AK58, wild emmer, and durum wheat. Dashed blue lines indicate collinearity between the 10 exons of *Q-5A* (ancestral) and the 7 exons of *Q-5B-hap3*. CACTA insertions in the last intron of *Q-5B-hap2* and *hap3* are indicated by orange lines.

**(B)** Gene CNV and co-located TEs for the photoperiod response gene *Ppd1*.

**(C)** Chromatin accessibility variation in the *Ppd1* promoter caused by a TE insertion and a deletion. This genome browser view displays signals for methylated histone sites and DNase I hypersensitive sites in ALB/78 (*A. tauschii*, DD), AK58, and CS for *Ppd-A1* and *Ppd-D1*. Read counts were normalized to counts per million and scaled; open chromatin signals are indicated by arrows. The absence of an open chromatin signal in the *Ppd-D1* promoter of AK58 indicates a TE deletion. Three motifs (blue bars; from left to right: G box CACGTG, G box CACGTC, and Sp1 GGGCGG) are shown in the open promoter region of *Ppd-A1* in AK58. A 100-bp conserved region (yellow rectangle) is located within the open region (red rectangle).

the short arms of the homoeologous group 3 chromosomes in the AK58 genome, with 3 TDs each of *Btr1* and *Btr2* in subA, 4 *Btr1* and 5 *Btr2* copies in subB, and 3 *Btr1* and 7 *Btr2* copies in subD, reflecting CNV among these homoeologous genes (Supplemental Figure 34A). Comparative analysis revealed that BR gene CNV among subgenomes primarily originates from differences in copy number between their diploid progenitors (Supplemental Figure 34B). In addition, the genes *TaBtr2\_1-A* and *TaBtr1\_1-A* may have originated during the domestication process from *T. urartu* to *Triticum monococcum*, whereas *TaBtr1\_4-B* may have originated during the development of the hexaploid cultivar AK58 from the hexaploid landrace CS, and *TaBtr2\_2-B* may have originated during the polyploidization event involving *A. speltoides* that produced wild emmer (Supplemental Table 15). TEs, primarily DNA/CACTA, LTR/Gypsy, and DNA/Mutator elements, were identified around *Btr1* and *Btr2*. These TEs likely induced *Btr* gene duplication through CACTA element capture or mispairing during meiosis. The presence of multiple *Btr* copies may reinforce BR in wild wheat, which is advantageous for wild species. Because wheat domestication primarily involves subA and subB, we analyzed the expression of *Btr* genes from these subgenomes. Notably, all of them exhibited no or very weak expression in spike tissues, whereas *Btr1-A* (*TraesAK58CH3A01G041100LC*) and *Btr1-B* (*TraesAK58CH3B01G123800*) were highly expressed in wild emmer (Supplemental Table 16), consistent with the loss of their function during domestication. We further searched for the fifth *Btr1-B* homolog (*TraesAK58CH3B01G123800*) in the available wheat genomes and identified intact LTR or solo-LTR Gypsy insertions around position 538 in the coding region, resulting in gene length variation ranging from 3786 bp in the wheat cultivar

Lancer to 14 646 bp in SY\_Mattis (Supplemental Table 17). Notably, the Gypsy insertions were present in cultivated wheat accessions but absent in wild emmer wheat. No expression was detected in homologs from cultivars such as AK58 and CS, whereas high expression levels were observed in the lemma and glume of wild emmer. Gypsy elements are associated with DNA methylation, which typically suppresses gene expression. This gene was highly methylated in the AK58 genome (Supplemental Figure 34C), consistent with the above hypothesis. These results suggest that TE insertions may be associated with wheat domestication.

The Q gene is considered important in the domestication of common wheat and is associated with several domestication-related traits, including free-threshing character, glume shape, spike length, and floral transition time (Faris et al., 2003; Simons et al., 2006; Zhang et al., 2011). Both coding sequence variations and EVs were identified in its AK58 homoeologs. The V329I mutation has been proposed to underlie the phenotypic differences between the Q and q alleles (Simons et al., 2006). Among the three Q homoeologs in the AK58 genome, this mutation was observed in both *Q-5A* and *Q-5D* (Supplemental Figure 35), whereas *Q-5B* was truncated due to a 531-bp deletion. Comparison of protein sequences from wheat progenitors revealed that the V329I mutation in *Q-5A* and *Q-5D* was present in the diploid progenitors, whereas the *Q-5B* truncation occurred after the domestication of allotetraploid wheat. We also observed a Gypsy insertion into intron 3 in the Julius and Rena varieties (Figure 8A; Supplemental Table 18), which may be associated with the exon deletion. In the AK58 genome, *Q-5A* displayed

dominant expression in spikes (Supplemental Figure 36A and 36B) and demonstrated significantly stronger co-expression intensity with its partner genes compared to *Q-5B* and *Q-5D* (Supplemental Figure 36C). Furthermore, almost no co-expressed genes were shared between *Q-5D* and the other two homoeologs. These results are consistent with the previously described roles of the three *Q* homoeologs: hyperfunctionalization of *Q-5A*, truncation of *Q-5B*, and sub-functionalization of *Q-5D*. Previous studies have reported that among the three *Q* homoeologs, only *Q-5A* is functionally relevant for domestication and other agronomic traits. In the present study, a homoeologous locus-based genome-wide association study (HGWAS) demonstrated that diversity among homoeoallele haplotypes resulting from homoeologous divergence was associated with variation in all three related agronomic traits (Supplemental Table 19). For example, the homoeoallele haplotype associated with the highest grain number per spike, CCA (i.e., CS alleles for subA and subB, and AK58 for subD), produced 67.8 grains per spike in the AK58/CS F2 population—26.9% higher than in AK58 (53.4 grains per spike)—which suggests a functional effect of homoeologous diversity. These findings indicate that *Q* continues to contribute to wheat improvement after domestication.

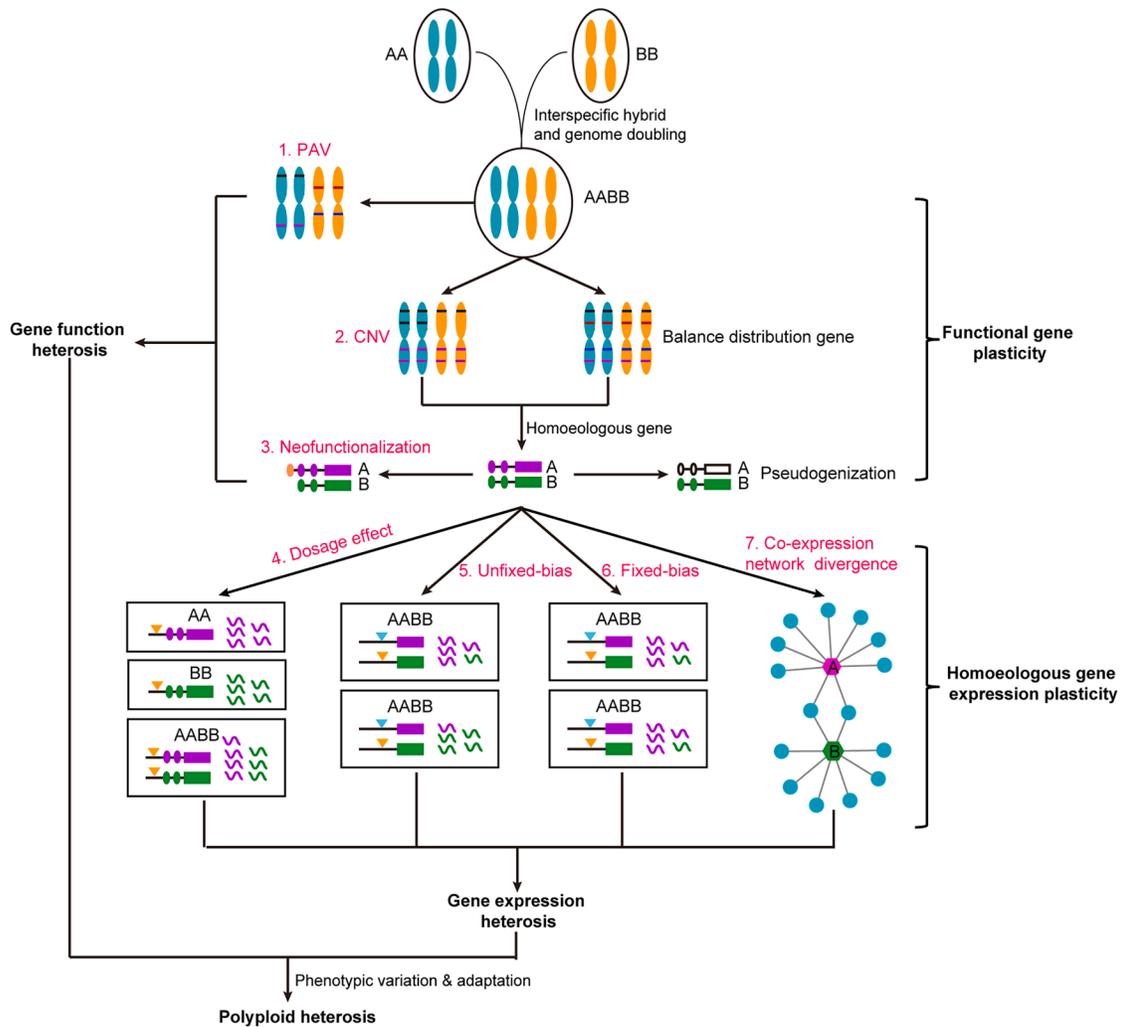
*Ppd1*, a gene from the pseudo-response regulator family, regulates multiple traits in wheat, including heading date, plant height, and grain weight. Photoperiod insensitivity enables wheat to adapt to a wide range of environments and contributed to the “Green Revolution” (Borlaug, 1983). This trait is associated with increased expression of *Ppd-B1* and *Ppd-D1*. *Ppd-B1* expression is associated with CNV (Wilhelm et al., 2009) and methylation in its promoter region (Sun et al., 2014), whereas *Ppd-D1* carries a 2089-bp deletion in the promoter region (Wilhelm et al., 2009; Guo et al., 2010). In the AK58 genome, we found that *Ppd-B1* was located within a 34-kb tandem segment duplication comprising three copies, each containing 3–4 TEs (Figure 8B). The duplication breakpoint was identified in an overlapping Gypsy element, which may have caused the TD through mismatched chromosome pairing during meiosis. The methylated region of the promoter was also located in the TE region (which contains a DNA TE and an unclassified TE) (Sun et al., 2014). Three DNA motifs associated with photoperiod sensitivity were identified in the deleted region of the *Ppd-D1* promoter in AK58 through comparison with the photoperiod-sensitive landrace CS (Figure 8C). Because both the deletion and methylation are associated with higher expression and occur within the promoter region, this region may contain a binding site for a transcriptional repressor. Therefore, we combined allele and homoeoallele comparisons to fine-map the loci associated with functional variation. A transcription factor binding site was mapped to a 100-bp segment of the promoter region, which included one of the 3 DNA motifs (Supplemental Figure 37A). This region was also found to harbor a “CACGTC” G box (Supplemental Figure 37B). The TEs in this region may have affected *Ppd1* expression by inducing deletions or promoting methylation of regulatory binding sites. HGWAS analysis indicated the functional importance of *Ppd1* homoeologous diversity across six related agronomic traits (Supplemental Table 20). For example, flag leaf length (FLL), an important agronomic trait for wheat improvement, is shorter in modern varieties than in landraces. We identified an elite homoeoallele haplotype, CAC, composed of *Ppd-A* and *Ppd-D* from CS and

*Ppd-B* from AK58, with an FLL of 17.7 cm, which is 4.5% shorter than in AK58 and 11.2% shorter than in CS, suggesting the potential of homoeologous variation for further wheat improvement.

Homoeologous divergence in both coding sequence and expression levels was also observed in the improvement gene *Rht1*. Among the *Rht1* homoeologs, *Rht-D1* encodes a mutant protein with a deletion of 36 amino acids in its DELLA domain (Supplemental Figure 38A), which prevents protein degradation and leads to constitutive activation of the GA signaling repressor, resulting in a dwarf phenotype (Dill et al., 2001). Compared to *Rht-A1* and *Rht-B1*, *Rht-D1* was predominantly expressed in stems and showed significantly higher correlation with several rice orthologs of internode architecture genes, including *OsCYP714B1-4A*, *OsSD1/GA20ox2-3B*, and *OSH15-2D* (Supplemental Figure 38B). The *Rht1* gene is associated with different TE distributions among the three subgenomes, and these TEs are closely associated with DNA methylation (Supplemental Figure 38C), which may contribute to the expression divergence of *Rht1* among the subgenomes. HGWAS analysis of the AK58/CS F2 population demonstrated the functional importance of *Rht1* homoeologous diversity for three agronomic traits: plant height (PH), flag leaf thickness (FLT), and grain width (Supplemental Table 20). FLT is a key architectural trait for wheat improvement, with modern varieties having thicker flag leaves than landraces. We identified an elite homoeoallelic haplotype, CAA, composed of *Rht1-A* from CS and *Rht1-B* and *Rht1-D* from AK58, associated with a flag leaf thickness of 0.23 mm, which is 12.1% greater than that of CS. These results not only demonstrate the functional importance of homoeologous divergence but also reveal an additional role for *Rht1*, the Green Revolution gene, in increasing FLT in addition to reducing plant height.

## DISCUSSION

The generation of heterosis is one of the most widely used strategies to enhance crop yield. As a double interspecific hybrid, hexaploid wheat exhibits extensive polyploid heterosis, also known as polyploid advantage or polyploid vigor (Dubcovsky and Dvorak, 2007). A comparison of global yields for diploid rice, polyploid wheat, and rapeseed from 1961 to 2018 showed that wheat and rapeseed yields increased by 1.5-fold and 1.9-fold, respectively, compared with only a 1.1-fold increase for rice (FAO, <http://www.fao.org/faostat/en/#data/QC>), which suggests greater yield potential in polyploid crops than in diploid crops. The success of polyploid plants lies mainly in the plasticity of their genomes, which results from their expanded genomic resources (Adams and Wendel, 2005; Comai, 2005; Dubcovsky and Dvorak, 2007; Chen, 2010). In this study, we found that most homoeologs exhibit similar functions but differ in expression levels and subfunctionalization. This suggests that homoeologs in polyploids behave analogously to alleles in heterozygous diploids; we therefore termed them homoeoalleles. Compared with allelic variation in diploids, homoeoallelic variation in polyploids is much greater (arithmetic progression vs. geometric progression). In addition, polyploid genomes are expected to contain more mutations than related diploid genomes (Pigiucci, 2007; Arrigo and Barker, 2012), which may enable the expression of advantageous homoeoalleles at critical



**Figure 9. Putative mechanisms for polyploid heterosis based on subgenome gene divergence.**

Allopolyploids are generally considered to be formed through the combined processes of interspecific hybridization and genome doubling, whereas autopolyploids arise via a single individual or crossing between individuals from genetically distinct lineages within the species.

spatiotemporal stages, thereby refining gene functions through long-term natural and artificial selection.

It has been shown that heterogeneity at a few genetic loci is responsible for heterosis in diploid hybrids (Krieger et al., 2010; Singh et al., 2013; Huang et al., 2016). Through association analysis of over 20 traits in a population derived from a cross between AK58 and CS, we previously found that HGWAS is more effective than single homoeolog-based GWAS in identifying the genetic basis of agronomic traits in polyploid wheat (Jia et al., 2023). Notably, elite homoeologous haplotypes, formed by combinations of subgenomic homoeologs of the associated loci, were identified in both parents and progeny, and many substantially improve wheat yield and related traits. These results suggest that, in polyploids, favorable alleles originate not only from different gene loci but also from homoeologous genes in different subgenomes. Further genetic and functional studies of homoeoallelic diversity and variation will clarify how they interact across subgenomes to confer polyploid heterosis. Overall, our results demonstrate that divergence among homoeologous genes contributes to the polyploid advantage

via increases in functional gene diversity and expression plasticity (Figure 9).

TEs account for 85% of the wheat genome and are important in generating homoeoallelic variation. In this study, we thoroughly examined the patterns of TE distribution across subgenomes and within or around functional genes in the polyploid wheat cultivar AK58. We found that TEs are important mediators of homoeologous gene divergence. Accordingly, the progressive accumulation of new TE insertions in polyploid wheat induced significant changes in genome structure and promoted divergence of the three subgenomes and their homoeologous genes. We found that genome-wide TE insertions into gene flanking regions and/or gene bodies cause CNV, structural variation, methylation variation, and expression variation. The ready-to-use *cis*-regulatory elements within TEs evidently influence gene expression, with important implications for wheat domestication and genetic improvement. The generation of diversity for subsequent selection is an essential principle in plant breeding. Almost all known agronomically important genes are associated with TEs. TEs are abundant, mobile, diverse, and subject to rapid

turnover; they are the primary contributors to genomic and functional diversity in polyploid wheat. The diploidization and divergence of homoeologs in polyploids form the basis of polyploid evolution and are crucial contributors to polyploid heterosis. Our findings show that TEs exhibit biased distribution across the three subgenomes and have played a significant role in homoeolog diploidization and subgenome divergence.

TEs represent a dynamic reservoir of genomic diversity. Inducing TE activity may generate elite alleles for wheat breeding, enabling the novel breeding strategy called breeding by induced TE activity (BITA). TE activity can be induced through tissue culture and demethylation (Orlowska et al., 2016; Song and Cao, 2017). In transformation studies, some mutant plants have been found to carry a target allele or gene without showing the desired agronomic traits. These mutants resulted from TE insertions or deletions induced during tissue culture as part of the transformation process (Orlowska et al., 2016; Song and Cao, 2017). Some new varieties bred from crosses between wheat and related species showed desirable agronomic traits that were absent in the original wheat parent, despite the absence of detectable non-wheat chromosomes or chromosome fragments in the resulting varieties. Some of these traits were likely induced by TE mutations during tissue culture or embryo rescue.

TE activity is generally suppressed by methylation in the host genome. Recently developed demethylation methods (Lei et al., 2015; Zhang et al., 2018) can be used in BITA. Environmental factors, especially abiotic stress, can also induce TE activity (Ngezahayo et al., 2009; Dubin et al., 2018). TE insertion mutants with desirable traits can be selected under these stress conditions. We previously identified active TEs (Wang et al., 2022b) that could be used in BITA. Because TE-induced mutations are random and typically produce only a few mutants with desirable agronomic traits, effective selection may require screening large populations across multiple environments.

## METHODS

### RNA-seq and expression profiling

Three classes of transcriptome datasets were used to assess dynamic gene expression. (1) Forty-two samples from AK58 across various tissues, developmental stages, and abiotic stress conditions (detailed in Supplemental Table 7) were collected for RNA-seq. With 3 biological replicates per sample, 126 RNA-seq libraries were generated. (2) Fourteen samples from seven tissues/organs of AK58 and CS (detailed in Supplemental Table 5) were collected for RNA-seq. With 3 biological replicates per sample, 42 RNA-seq libraries were generated. (3) Fifty-seven samples from a newly synthesized wheat allohexaploid and its donor tetraploid and diploid progenitors (detailed in Supplemental Table 8) were collected for RNA-seq. With 1–3 replicates per sample, 105 RNA-seq libraries were generated. The overall dataset contains 273 RNA-seq libraries for wheat and its progenitors, covering 6 organs, 11 developmental stages, and 6 abiotic stress conditions. For each replicate, at least five plants were collected and pooled for RNA extraction.

Messenger RNA from all samples was used to construct paired-end RNA-seq libraries, which were sequenced on an Illumina HiSeq 2500 platform at Novogene Bioinformatics Technology (Beijing, China). RNA-seq reads were mapped to the AK58 genome using HISAT2 with default parameters (Pertea et al., 2016). Read pair alignments with conflicts at a single locus or mapping to multiple loci were filtered out. Concordant paired-end alignments were retained and used to determine gene expression levels

using StringTie (v.1.3.3) (Pertea et al., 2016). Genes with an FPKM value greater than one in any sample were selected for further analysis.

### Analysis of sample relatedness and identification of compartment-specific genes

To facilitate the graphical interpretation of sample relatedness, FPKM values for the expressed genes were reduced to three dimensions using principal-component analysis with the `prcomp()` function in R. Genes expressed specifically in each genomic compartment were identified using a scoring algorithm, as previously described (Yanai et al., 2005). Genes designated as compartment-specific should also meet the criteria that they have an FPKM  $\geq 5$  in at least one sample.

### Co-expression network analysis and functional annotation

All expressed HC genes were used to construct the co-expression network using the WGCNA R package (Langfelder and Horvath, 2008). A soft-thresholding power of 5 was used because it was the lowest power at which the scale-free topology fit index reached 0.9. The `blockwiseModules()` function in WGCNA was used to construct 2 blocks, with the maximum block size set to 46 000 genes. Other parameters were set as follows: `maxPOutliers = 0.05`, `TOMType = "unsigned"`, `mergeCutHeight = 0.15`, and `minModuleSize = 30`. The most highly correlated genes, identified using the `signedKME()` function, were considered central to their module. Co-expression networks for the specified genes were visualized using Cytoscape (v.3.5.1) (Daron et al., 2014). Functional information for rice genes was obtained from the `funRiceGenes` database (<http://funricegenes.ncpgr.cn/>).

### Homoeologous gene identification in AK58 and related wheat genomes

All annotated HC coding sequences from different subgenomes were subjected to an all-versus-all BLAST search (NCBI-BLAST v.2.2.27) with the parameters `E < 1 × 10-10` and `-max_target_seq 1`. Syntenic blocks among the three subgenomes were identified using MCScanX (-b 2, -s 5) (Wang et al., 2012). The set of AK58 HC genes was separated by subgenomic origin (A, B, and D) and used to infer gene families and homoeologous gene groups via OrthoMCL (v.2.0) with default parameters. Homoeologs, which were defined as orthologs among the subgenomes in OrthoMCL, were treated as distinct taxa. Syntenic homoeologous gene groups were defined as homoeologs that were part of a conserved collinear block among the subgenomes.

In the AK58 genome, a total of 119 448 genes were annotated, of which 63 843 (53.4%; 53 349 syntenic triads and 10 494 non-syntenic triads) were present as triad homoeologs, i.e., having 1 copy in each of the 3 subgenomes (A:B:D = 1:1:1). Among the homoeologous genes, 21 232 (17.8%) had at least 1 subgenome paralog, defined as more than 1 gene copy in a single subgenome (A:B:D = 1:1:N, 1:N:1, N:1:1, or another ratio). There were 10 192 genes (8.5%) present as dyads, i.e., with the loss of one homoeolog (A:B:D = 1:1:0, 1:0:1, or 0:1:1), with similar levels across the A, B, and D subgenomes. The remaining 22 051 genes (18.5%) were singletons, i.e., with no homoeologs in other subgenomes, and appeared to be subgenome-specific paralogs (A:B:D = 1:0:0, 0:1:0, or 0:0:1). A total of 2130 genes were located on unanchored contigs.

All HC genes derived from AK58, CS, wild emmer (the allotetraploid ancestor, AABB), and *A. tauschii* (the D genome ancestor) were used to infer orthologous relationships using OrthoMCL (v.2.0) with default parameters. These relationships were used to assess whether the absence of homoeologous loci in AK58 was already present in the ancestors of hexaploid wheat. Using protein-guided DNA alignments, we estimated non-synonymous ( $K_a$ ) and synonymous ( $K_s$ ) substitution rates for pairwise comparisons of homoeologs in each orthologous group using PAML with the F3X4 model (Yang, 2007). Pfam Scan ([www.ebi.ac.uk/Tools/pfam/pfamscan/](http://www.ebi.ac.uk/Tools/pfam/pfamscan/)) was used to identify putative protein functional domains by searching against the Pfam HMM library (v.32.0).

### Expression divergence analysis across homoeologous genes

In the current study, three methods were used to evaluate biased expression among homoeologous genes. (1) For each group of homoeologous genes, locus-wise FPKM values were transformed as  $\log_2(\text{FPKM}+0.01)$  and used to test for significant deviations in expression levels among subgenomes. Biased expression was defined as the presence of homoeologs with significantly different expression levels in any of the three possible pairwise comparisons (subA vs. subB, subA vs. subD, and subB vs. subD; Fisher's exact test, BH-corrected  $p < 0.05$ , fold change  $\geq 1.5$ ) (Liu et al., 2015). (2) Based on co-expression analysis, all expressed genes were classified into 84 co-expression modules. Homoeologous genes assigned to different co-expression modules were defined as exhibiting co-expression module divergence. (3) The top 100 co-expressed genes, based on co-expression weight values, were identified for each gene. The proportion of shared co-expressed genes was calculated for each of the three possible pairwise homoeolog comparisons (subA vs. subB, subA vs. subD, and subB vs. subD). Co-expression gene divergence was defined as the presence of homoeologous genes with a proportion of shared co-expressed genes less than 30%. Functional analysis of homoeologs with biased expression in each module was performed using GO, PO, and TO terms inferred from the AK58 gene annotation, and enrichment analysis was performed using the R package clusterProfiler (Yu et al., 2012).

### Analysis of accelerated evolution in physically linked genes and functional categories

Across subgenomes, extreme outliers were identified by comparing the mean Ka/Ks ratios for sliding windows of 10 homoeologs with the average ratio across all homoeologs, using a metric based on the binomial test (Mikkelsen et al., 2005). The number of observed outliers (test statistic  $< 0.001$ ) was then compared with the expected distribution, generating a set of randomly selected homoeologs. Regions undergoing accelerated selection were subsequently grouped into clusters based on the physical positions of the loci. Similarity analysis was carried out for each GO, PO, and TO category containing at least 20 genes to identify fast- or slow-evolving categories across subgenomes.

### Annotation and analysis of TEs, pseudogenes, and gene duplication patterns

The genome sequence and annotation for *A. tauschii* were downloaded from NCBI (BioProject ID: PRJNA182898). Genomic sequences and corresponding annotations for the remaining plant species (*T. aestivum* cv. CS, *T. turgidum* var. *dicoccoides*, and *T. turgidum* ssp. *durum*) were obtained from the EnsemblPlants database. TE and pseudogene annotations were performed as described in our previously published genomics article (Jia et al., 2023). In brief, each genome sequence was investigated for TE content using RepeatMasker with the ClariTERep database (<https://github.com/jdaron/CLARI-TE>), which catalogs repetitive elements in the wheat genome. The raw TE similarity search results were refined using the CLARI-TE Perl script, as previously described (Daron et al., 2014). Pseudogene annotation was performed using a homology-based approach: BLAST was used to compare the TE-masked sequences with all HC transcripts from the AK58 genome. Pseudogenes were defined as models with a canonical gene structure but less than 70% similarity to the full length of the best BLAST hit, as well as models without start and/or stop codons. Gene duplication patterns were identified using DupGen\_finder (Qiao et al., 2019).

### Analysis of TEs in the vicinity of genes and pseudogenes

TEs co-located with genes were identified by overlapping the feature coordinates of TEs with those of genes in each genome. To calculate TE density within each gene body and adjacent regions, we divided the gene body into 100 equal segments and calculated the proportion of TE bases within each segment. The mean density was then computed for each gene category. For the flanking regions (2 kb upstream and downstream of the gene), TE density was calculated using a sliding window approach with a

window size of 100 bp. The mean density was subsequently determined for each window across each categorized gene. Using a previously reported method (Daron et al., 2014), we identified potential CACTA-mediated capture events by searching for genes flanked by two elements from the CACTA family.

### Analysis of DNA methylation, MNase-seq, and chromatin immunoprecipitation sequencing data

The integration of gene expression profiles from RNA-seq, genome-wide DNA methylation data, and chromatin epigenomic states can provide a multi-dimensional perspective on the regulation of gene expression. Therefore, to investigate the role of TEs in mediating homoeologous gene divergence, we integrated DNA bisulfite sequencing, MNase-seq, and chromatin immunoprecipitation sequencing data from the leaf and root tissues of seedlings at the third leaf stage, as described in our previous studies (Jia et al., 2021, 2023). Cytosine methylation levels were determined using Bismark (v.0.16.1) (Krueger and Andrews, 2011), retaining only unique and unambiguous alignments. Methylation ratios for all CG, CHG, and CHH sites were calculated as the number of methylated cytosines divided by the number of reads covering each site. Genes were classified as associated with methylation if cytosines within the gene body or promoter region were methylated.

MNase-seq data were used to evaluate chromatin accessibility. After preprocessing with Trimmomatic v0.36 (Bolger et al., 2014), clean reads were aligned to the genome using Bowtie2 (v.2.3.4) (Langmead and Salzberg, 2012). Read coverage in 10-bp intervals was calculated using the Bedtools coverage function (Quinlan and Hall, 2010). Difference values were obtained by subtracting the mean normalized depth of the heavy digest replicates from that of the light digest replicates. Positive values were defined as MNase hypersensitive, and a Bayes factor criterion was used to classify significantly hypersensitive regions, as previously described (Rodgers-Melnick et al., 2016). Methylation, MNase, and histone signals were annotated using the gene feature coordinates and 2-kb flanking regions. Next, the Euclidean distances between each triad and the three vertices of the ternary plot were determined based on the relative modification levels of the homoeologs, as previously described (Ramirez-Gonzalez et al., 2018). Finally, all triads were categorized into seven methylation patterns (i.e., subgenome A, B, or D dominant; subgenome A, B, or D suppressed; and similar methylation abundance across subgenomes) based on the shortest Euclidean distance. *cis*-acting regulatory elements in open regions were identified using PlantCARE tools (Lescot et al., 2002).

### DATA AVAILABILITY

The data supporting the findings of this study are provided in the main text and supplemental materials. The whole-genome sequence and transcriptome data reported in this paper have been deposited in the Genome Warehouse of the National Genomics Data Center (National Genomics Data Center, 2020), Beijing Institute of Genomics (China National Center for Bioinformatics), Chinese Academy of Sciences, under accession number GWHANRF00000000 and are publicly accessible at <https://bigd.big.ac.cn/gwh>. Additional data are available from the corresponding author upon reasonable request.

### FUNDING

We are grateful to Prof. Yijing Zhang and Yilin Xie for valuable assistance with methylation data analysis. This work was supported by the National Natural Science Foundation of China (91940301, 32441002, 32370341, and 32201844), the Natural Science Basic Research Plan in Shaanxi Province of China (2022JQ-172), Zhongyuan Scholar Program (234000510002), the National Key Research and Development Program of China (2021YFD1200601-04), and the Hundred-Talent Program of

Zhejiang University. We also thank the High-Performance Computing (HPC) facility of the State Key Laboratory of Crop Stress Biology in Arid Areas and Northwest A&F University for providing computing resources.

## ACKNOWLEDGMENTS

No conflict of interest is declared.

## AUTHOR CONTRIBUTIONS

J.J. and L.W. conceived and supervised the research. P.D., T.L., and K.W. performed the bioinformatics analyses. K.W., L.G., and G.Z. prepared materials for RNA sequencing and assisted with data analysis. D.C., Z.D., K.Z., and Z.R. contributed research materials. P.D., J.J., W.D., L.W., W.J., and C.L. wrote and revised the article.

## SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: July 30, 2024

Revised: January 10, 2025

Accepted: April 14, 2025

Published: April 16, 2025

## REFERENCES

- Adams, K.L.** (2007). Evolution of duplicate gene expression in polyploid and hybrid plants. *J. Hered.* **98**:136–141. <https://doi.org/10.1093/jhered/esl061>.
- Adams, K.L., and Wendel, J.F.** (2005). Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **8**:135–141. <https://doi.org/10.1016/j.pbi.2005.01.001>.
- Adamski, N.M., Borrill, P., Brinton, J., Harrington, S.A., Marchal, C., Bentley, A.R., Bovill, W.D., Cattivelli, L., Cockram, J., Contreras-Moreira, B., et al.** (2020). A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat. *Elife* **9**:e55646. <https://doi.org/10.7554/eLife.55646>.
- Alfatih, A., Wu, J., Jan, S.U., Zhang, Z.S., Xia, J.Q., and Xiang, C.B.** (2020). Loss of rice PARAQUAT TOLERANCE 3 confers enhanced resistance to abiotic stresses and increases grain yield in field. *Plant Cell Environ.* **43**:2743–2754.
- Arrigo, N., and Barker, M.S.** (2012). Rarely successful polyploids and their legacy in plant genomes. *Curr. Opin. Plant Biol.* **15**:140–146. <https://doi.org/10.1016/j.pbi.2012.03.010>.
- Baduel, P., Quadrana, L., Hunter, B., Bomblies, K., and Colot, V.** (2019). Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat. Commun.* **10**:5818. <https://doi.org/10.1038/S41467-019-13730-0>.
- Bansal, P., Banga, S., and Banga, S.S.** (2012). Heterosis as Investigated in Terms of Polyploidy and Genetic Diversity Using Designed Brassica juncea Amphiploid and Its Progenitor Diploid Species. *PLoS One* **7**:e29607. <https://doi.org/10.1371/journal.pone.0029607>.
- Bolger, A.M., Lohse, M., and Usadel, B.** (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Borlaug, N.E.** (1983). Contributions of Conventional Plant-Breeding to Food-Production. *Science* **219**:689–693. <https://doi.org/10.1126/science.219.4585.689>.
- Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvak, Z., Levin, H.L., Macfarlan, T. S., et al.** (2018). Ten things you should know about transposable elements. *Genome Biol.* **19**:199. <https://doi.org/10.1186/S13059-018-1577-Z>.
- Chen, Z.J.** (2007). Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu. Rev. Plant Biol.* **58**:377–406. <https://doi.org/10.1146/annurev.arplant.58.032806.103835>.
- Chen, Z.J.** (2010). Molecular mechanisms of polyploidy and hybrid vigor. *Trends Plant Sci.* **15**:57–71. <https://doi.org/10.1016/j.tplants.2009.12.003>.
- Chen, Z.J.** (2013). Genomic and epigenetic insights into the molecular bases of heterosis. *Nat. Rev. Genet.* **14**:471–482. <https://doi.org/10.1038/nrg3503>.
- Cheng, F., Wu, J., Cai, X., Liang, J., Freeling, M., and Wang, X.** (2018). Gene retention, fractionation and subgenome differences in polyploid plants. *Nat. Plants* **4**:258–268. <https://doi.org/10.1038/s41477-018-0136-7>.
- Chi, L.J., Saarela, U., Railo, A., Prunskaitė-Hyyryläinen, R., Skovorodkin, I., Anthony, S., Katsu, K., Liu, Y., Shan, J.D., Salgueiro, A.M., et al.** (2011). A Secreted BMP Antagonist, Cer1, Fine Tunes the Spatial Organization of the Ureteric Bud Tree during Mouse Kidney Development. *PLoS One* **6**:e27676. <https://doi.org/10.1371/journal.pone.0027676>.
- Coghlan, A., Eichler, E.E., Oliver, S.G., Paterson, A.H., and Stein, L.** (2005). Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet.* **21**:673–682. <https://doi.org/10.1016/j.tig.2005.09.009>.
- Comai, L.** (2005). The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**:836–846. <https://doi.org/10.1038/nrg1711>.
- Daccord, N., Celton, J.M., Linsmith, G., Becker, C., Choisne, N., Schijlen, E., van de Geest, H., Bianco, L., Micheletti, D., Velasco, R., et al.** (2017). High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* **49**:1099–1106. <https://doi.org/10.1038/ng.3886>.
- Daron, J., Glover, N., Pingault, L., Theil, S., Jamilloux, V., Paux, E., Barbe, V., Mangenot, S., Alberti, A., Wincker, P., et al.** (2014). Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biol.* **15**:546. <https://doi.org/10.1186/s13059-014-0546-4>.
- Dill, A., Jung, H.S., and Sun, T.P.** (2001). The DELLA motif is essential for gibberellin-induced degradation of RGA. *Proc. Natl. Acad. Sci. USA* **98**:14162–14167. <https://doi.org/10.1073/pnas.251534098>.
- Dubcovsky, J., and Dvorak, J.** (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science* **316**:1862–1866. <https://doi.org/10.1126/science.1143986>.
- Dubin, M.J., Mittelsten Scheid, O., and Becker, C.** (2018). Transposons: a blessing curse. *Curr. Opin. Plant Biol.* **42**:23–29. <https://doi.org/10.1016/j.pbi.2018.01.003>.
- Edwards, R., Dixon, D.P., and Walbot, V.** (2000). Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends Plant Sci.* **5**:193–198.
- Faris, J.D., Fellers, J.P., Brooks, S.A., and Gill, B.S.** (2003). A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. *Genetics* **164**:311–321. <https://doi.org/10.1093/genetics/164.1.311>.
- Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I., and Moore, G.** (2006). Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* **439**:749–752.
- Guo, Z., Song, Y., Zhou, R., Ren, Z., and Jia, J.** (2010). Discovery, evaluation and distribution of haplotypes of the wheat Ppd-D1 gene. *New Phytol.* **185**:841–851. <https://doi.org/10.1111/j.1469-8137.2009.03099.x>.
- Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., Zhao, Y., Li, W., Cheng, B., Xia, J., et al.** (2016). Genomic architecture of heterosis for yield traits in rice. *Nature* **537**:629–633. <https://doi.org/10.1038/nature19760>.

- International Wheat Genome Sequencing Consortium IWGSC.** (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**:eaar7191. <https://doi.org/10.1126/science.aar7191>.
- Jia, J.Z., Xie, Y.L., Cheng, J.F., Kong, C.Z., Wang, M.Y., Gao, L.F., Zhao, F., Guo, J.Y., Wang, K., Li, G.W., et al.** (2021). Homology-mediated inter-chromosomal interactions in hexaploid wheat lead to specific subgenome territories following polyploidization and introgression. *Genome Biol.* **22**:26. <https://doi.org/10.1186/s13059-020-02225-7>.
- Jia, J., Zhao, G., Li, D., Wang, K., Kong, C., Deng, P., Yan, X., Zhang, X., Lu, Z., Xu, S., et al.** (2023). Genome resources for the elite bread wheat cultivar Aikang 58 and mining of elite homeologous haplotypes for accelerating wheat improvement. *Mol. Plant* **16**:1893–1910. <https://doi.org/10.1016/j.molp.2023.10.015>.
- Krieger, U., Lippman, Z.B., and Zamir, D.** (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nat. Genet.* **42**:459–463. <https://doi.org/10.1038/ng.550>.
- Krueger, F., and Andrews, S.R.** (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**:1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>.
- Labrou, N.E., Papageorgiou, A.C., Pavli, O., and Flemetakis, E.** (2015). Plant GSTome: structure and functional role in xenome network and plant stress response. *Curr. Opin. Biotechnol.* **32**:186–194. <https://doi.org/10.1016/j.copbio.2014.12.024>.
- Langfelder, P., and Horvath, S.** (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* **9**:559. <https://doi.org/10.1186/1471-2105-9-559>.
- Langmead, B., and Salzberg, S.L.** (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**:357–359. <https://doi.org/10.1038/nmeth.1923>.
- Lei, M., Zhang, H., Julian, R., Tang, K., Xie, S., and Zhu, J.K.** (2015). Regulatory link between DNA methylation and active demethylation in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **112**:3553–3557. <https://doi.org/10.1073/pnas.1502279112>.
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., and Rombauts, S.** (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**:325–327. <https://doi.org/10.1093/Nar/30.1.325>.
- Li, Z.J., Wang, M.Y., Lin, K.D., Xie, Y.L., Guo, J.Y., Ye, L.H., Zhuang, Y. L., Teng, W., Ran, X.J., Tong, Y.P., et al.** (2019). The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. *Genome Biol.* **20**:139. <https://doi.org/10.1186/S13059-019-1746-8>.
- Ling, H.Q., Ma, B., Shi, X., Liu, H., Dong, L., Sun, H., Cao, Y., Gao, Q., Zheng, S., Li, Y., et al.** (2018). Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature* **557**:424–428. <https://doi.org/10.1038/s41586-018-0108-0>.
- Liu, Z., Xin, M., Qin, J., Peng, H., Ni, Z., Yao, Y., and Sun, Q.** (2015). Temporal transcriptome profiling reveals expression partitioning of homeologous genes contributing to heat and drought acclimation in wheat (*Triticum aestivum* L.). *BMC Plant Biol.* **15**:152. <https://doi.org/10.1186/s12870-015-0511-8>.
- Mayer, K.F.X., Rogers, J., Dolezel, J., Pozniak, C., Eversole, K., Feuillet, C., Gill, B., Friebe, B., Lukaszewski, A.J., Sourdille, P., et al.** (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* **345**:1251788. <https://doi.org/10.1126/Science.1251788>.
- Mikkelsen, T.S., Hillier, L.W., Eichler, E.E., Zody, M.C., Jaffe, D.B., Yang, S.P., Enard, W., Hellmann, I., Lindblad-Toh, K., Altheide, T. K., et al.** (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**:69–87. <https://doi.org/10.1038/nature04072>.
- National Genomics Data Center, M.; Partners** (2020). Database Resources of the National Genomics Data Center in 2020. *Nucleic acids research* **48**:D24–D33. <https://doi.org/10.1093/nar/gkz913>.
- Ngezahayo, F., Xu, C., Wang, H., Jiang, L., Pang, J., and Liu, B.** (2009). Tissue culture-induced transpositional activity of mPing is correlated with cytosine methylation in rice. *BMC Plant Biol.* **9**:91. <https://doi.org/10.1186/1471-2229-9-91>.
- Ni, Z., Kim, E.D., Ha, M., Lackey, E., Liu, J., Zhang, Y., Sun, Q., and Chen, Z.J.** (2009). Altered circadian rhythms regulate growth vigour in hybrids and allopolyploids. *Nature* **457**:327–331. <https://doi.org/10.1038/nature07523>.
- Orlowska, R., Machczynska, J., Oleszczuk, S., Zimny, J., and Bednarek, P.T.** (2016). DNA methylation changes and TE activity induced in tissue cultures of barley (*Hordeum vulgare* L.). *J. Biol. Res-Thessalon* **23**:19. <https://doi.org/10.1186/s40709-016-0056-5>.
- Parisod, C., Salmon, A., Zerjal, T., Tenaillon, M., Grandbastien, M.A., and Ainouche, M.** (2009). Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytol.* **184**:1003–1015. <https://doi.org/10.1111/j.1469-8137.2009.03029.x>.
- Park, H.-Y., Kang, I.S., Han, J.-S., Lee, C.-H., An, G., and Moon, Y.-H.** (2009). OsDEG10 encoding a small RNA-binding protein is involved in abiotic stress signaling. *Biochem. Biophys. Res. Commun.* **380**:597–602.
- Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., et al.** (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* **457**:551–556. <https://doi.org/10.1038/nature07723>.
- Perteau, M., Kim, D., Perteau, G.M., Leek, J.T., and Salzberg, S.L.** (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**:1650–1667. <https://doi.org/10.1038/nprot.2016.095>.
- Pfeifer, M., Kugler, K.G., Sandve, S.R., Zhan, B., Rudi, H., Hvidsten, T. R., International Wheat Genome Sequencing Consortium, Mayer, K.F.X., and Olsen, O.A.** (2014). Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**:1250091. <https://doi.org/10.1126/Science.1250091>.
- Pigliucci, M.** (2007). The origins of genome architecture. *Science* **317**:1172–1173. <https://doi.org/10.1126/science.1146047>.
- Pourkheirandish, M., Dai, F., Sakuma, S., Kanamori, H., Distelfeld, A., Willcox, G., Kawahara, T., Matsumoto, T., Kilian, B., and Komatsuda, T.** (2017). On the Origin of the Non-brittle Rachis Trait of Domesticated Einkorn Wheat. *Front. Plant Sci.* **8**:2031. <https://doi.org/10.3389/Fpls.2017.02031>.
- Qiao, X., Li, Q.H., Yin, H., Qi, K.J., Li, L.T., Wang, R.Z., Zhang, S.L., and Paterson, A.H.** (2019). Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* **20**:38. <https://doi.org/10.1186/s13059-019-1650-2>.
- Quinlan, A.R., and Hall, I.M.** (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- Ramachandran, D., Mckain, M.R., Kellogg, E.A., and Hawkins, J.S.** (2020). Evolutionary Dynamics of Transposable Elements Following a Shared Polyploidization Event in the Tribe Andropogoneae. *G3 Genes|Genomes|Genetics* **10**:4387–4398. <https://doi.org/10.1534/g3.120.401596>.

- Ramirez-Gonzalez, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., van Ex, F., Pasha, A., et al. (2018). The transcriptional landscape of polyploid wheat. *Science* **361**:eaar6089. <https://doi.org/10.1126/science.aar6089>.
- Rodgers-Melnick, E., Vera, D.L., Bass, H.W., and Buckler, E.S. (2016). Open chromatin reveals the functional maize genome. *Proc. Natl. Acad. Sci. USA* **113**:E3177–E3184. <https://doi.org/10.1073/pnas.1525244113>.
- Ronfort, J. (1999). The mutation load under tetrasomic inheritance and its consequences for the evolution of the selfing rate in autotetraploid species. *Genet. Res.* **74**:31–42.
- Roulin, A., Auer, P.L., Libault, M., Schlueter, J., Farmer, A., May, G., Stacey, G., Doerge, R.W., and Jackson, S.A. (2013). The fate of duplicated genes in a polyploid plant genome. *Plant J.* **73**:143–153. <https://doi.org/10.1111/tpj.12026>.
- Shao, L., Xing, F., Xu, C., Zhang, Q., Che, J., Wang, X., Song, J., Li, X., Xiao, J., Chen, L.L., et al. (2019). Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl. Acad. Sci. USA* **116**:5653–5658. <https://doi.org/10.1073/pnas.1820513116>.
- Shi, X.L., Ng, D.W.K., Zhang, C.Q., Comai, L., Ye, W.X., and Chen, Z.J. (2012). Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in Arabidopsis allopolyploids. *Nat. Commun.* **3**:950. <https://doi.org/10.1038/Ncomms1954>.
- Simons, K.J., Fellers, J.P., Trick, H.N., Zhang, Z., Tai, Y.S., Gill, B.S., and Faris, J.D. (2006). Molecular characterization of the major wheat domestication gene Q. *Genetics* **172**:547–555. <https://doi.org/10.1534/genetics.105.044727>.
- Singh, R., Low, E.T.L., Ooi, L.C.L., Ong-Abdullah, M., Ting, N.C., Nagappan, J., Nookiah, R., Amiruddin, M.D., Rosli, R., Manaf, M.A.A., et al. (2013). The oil palm SHELL gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* **500**:340–344. <https://doi.org/10.1038/nature12356>.
- Song, Q.X., Lu, X., Li, Q.T., Chen, H., Hu, X.Y., Ma, B., Zhang, W.K., Chen, S.Y., and Zhang, J.S. (2013). Genome-Wide Analysis of DNA Methylation in Soybean. *Mol. Plant* **6**:1961–1974. <https://doi.org/10.1093/mp/sst123>.
- Song, X., and Cao, X. (2017). Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Curr. Opin. Plant Biol.* **36**:111–118. <https://doi.org/10.1016/j.pbi.2017.02.004>.
- Springer, N.M., Ying, K., Fu, Y., Ji, T.M., Yeh, C.T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., et al. (2009). Maize Inbreds Exhibit High Levels of Copy Number Variation (CNV) and Presence/Absence Variation (PAV) in Genome Content. *PLoS Genet.* **5**:e1000734. <https://doi.org/10.1371/journal.pgen.1000734>.
- Sun, H., Guo, Z., Gao, L., Zhao, G., Zhang, W., Zhou, R., Wu, Y., Wang, H., An, H., and Jia, J. (2014). DNA methylation pattern of Photoperiod-B1 is associated with photoperiod insensitivity in wheat (*Triticum aestivum*). *New Phytol.* **204**:682–692. <https://doi.org/10.1111/nph.12948>.
- Takahashi, S., Inagaki, Y., Satoh, H., Hoshino, A., and Iida, S. (1999). Capture of a genomic HMG domain sequence by the En/Spm-related transposable element Tpn1 in the Japanese morning glory. *Mol. Genet.* **261**:447–451. <https://doi.org/10.1007/s004380050987>.
- Van de Peer, Y., Fawcett, J.A., Proost, S., Sterck, L., and Vandepoele, K. (2009). The flowering world: a tale of duplications. *Trends Plant Sci.* **14**:680–688. <https://doi.org/10.1016/j.tplants.2009.09.001>.
- Vaucheret, H. (2008). Plant ARGONAUTES. *Trends Plant Sci.* **13**:350–358. <https://doi.org/10.1016/j.tplants.2008.04.007>.
- Vicient, C.M., and Casacuberta, J.M. (2017). Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* **120**:195–207. <https://doi.org/10.1093/aob/mcx078>.
- Wang, H., Sun, S., Ge, W., Zhao, L., Hou, B., Wang, K., Lyu, Z., Chen, L., Xu, S., Guo, J., et al. (2020). Horizontal gene transfer of Fhb7 from fungus underlies Fusarium head blight resistance in wheat. *Science* **368**:eaba5435. <https://doi.org/10.1126/science.aba5435>.
- Wang, M., Li, Z., Zhang, Y., Zhang, Y., Xie, Y., Ye, L., Zhuang, Y., Lin, K., Zhao, F., and Guo, J. (2021). An atlas of wheat epigenetic regulatory elements reveals subgenome-divergence in the regulation of development and stress responses. *Plant Cell* **33**:865–881.
- Wang, X.L., Yan, X.Q., Hu, Y.H., Qin, L.Y., Wang, D.W., Jia, J.Z., and Jiao, Y.N. (2022a). A recent burst of gene duplications in Triticeae. *Plant Commun.* **3**:100268. <https://doi.org/10.1016/J.Xplc.2021.100268>.
- Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.H., Jin, H., Marler, B., Guo, H., et al. (2012). MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**:e49. <https://doi.org/10.1093/nar/gkr1293>.
- Wang, Z., Zhao, G., Yang, Q., Gao, L., Liu, C., Ru, Z., Wang, D., Jia, J., and Cui, D. (2022b). Helitron and CACTA DNA transposons actively reshape the common wheat - AK58 genome. *Genomics* **114**:110288.
- Washburn, J.D., and Birchler, J.A. (2014). Polyploids as a "model system" for the study of heterosis. *Plant Reprod.* **27**:1–5. <https://doi.org/10.1007/s00497-013-0237-4>.
- Wei, N., Cronn, R., Liston, A., and Ashman, T.L. (2019). Functional trait divergence and trait plasticity confer polyploid advantage in heterogeneous environments. *New Phytol.* **221**:2286–2297. <https://doi.org/10.1111/nph.15508>.
- Wicker, T., Gundlach, H., Spannagl, M., Uauy, C., Borrill, P., Ramirez-Gonzalez, R.H., De Oliveira, R., Mayer, K.F.X., Paux, E., Choulet, F., et al. (2018). Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biol.* **19**:103. <https://doi.org/10.1186/S13059-018-1479-0>.
- Wilhelm, E.P., Turner, A.S., and Laurie, D.A. (2009). Photoperiod insensitive Ppd-A1a mutations in tetraploid wheat (*Triticum durum* Desf.). *Theor. Appl. Genet.* **118**:285–294. <https://doi.org/10.1007/s00122-008-0898-9>.
- Xiao, J., Liu, B., Yao, Y., Guo, Z., Jia, H., Kong, L., Zhang, A., Ma, W., Ni, Z., Xu, S., et al. (2022). Wheat genomic study for genetic improvement of traits in China. *Sci. China Life Sci.* **65**:1718–1775. <https://doi.org/10.1007/s11427-022-2178-7>.
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., et al. (2005). Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics* **21**:650–659. <https://doi.org/10.1093/bioinformatics/bti042>.
- Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**:1586–1591. <https://doi.org/10.1093/molbev/msm088>.
- Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics-a Journal of Integrative Biology* **16**:284–287. <https://doi.org/10.1089/omi.2011.0118>.
- Zabala, G., and Vodkin, L.O. (2005). The wp mutation of Glycine max carries a gene-fragment-rich transposon of the CACTA superfamily. *Plant Cell* **17**:2619–2632. <https://doi.org/10.1105/tpc.105.033506>.

## Plant Communications

**Zhang, H., Lang, Z., and Zhu, J.K.** (2018). Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**:489–506. <https://doi.org/10.1038/s41580-018-0016-z>.

**Zhang, Z., Belcram, H., Gornicki, P., Charles, M., Just, J., Huneau, C., Magdelenat, G., Couloux, A., Samain, S., Gill, B.S., et al.** (2011). Duplication and partitioning in evolution and function of homoeologous Q loci governing domestication characters in polyploid wheat. *Proc. Natl. Acad. Sci. USA* **108**:18737–18742. <https://doi.org/10.1073/pnas.1110552108>.

## Subgenome divergence in hexaploid wheat

**Zhao, G., Zou, C., Li, K., Wang, K., Li, T., Gao, L., Zhang, X., Wang, H., Yang, Z., Liu, X., et al.** (2017). The *Aegilops tauschii* genome reveals multiple impacts of transposons. *Nat. Plants* **3**:946–955. <https://doi.org/10.1038/s41477-017-0067-8>.

**Zhu, T., Nevo, E., Sun, D., and Peng, J.** (2012). Phylogenetic analyses unravel the evolutionary history of NAC proteins in plants. *Evolution* **66**:1833–1848. <https://doi.org/10.1111/j.1558-5646.2011.01553.x>.