Plant Communications
Research article

CellPress
Partner Journal

Centromere repositioning and shifts in wheat evolution

Jing Zhao[1,3,8], Yilin Xie[2,7,8], Chuizheng Kong[1,8], Zefu Lu[1], Haiyan Jia[3], Zhengqiang Ma[3], Yijing Zhang[2], Dangqun Cui[4], Zhengang Ru[5], Yuquan Wang[5], Rudi Appels[6], Jizeng Jia[1,4,*] and Xueyong Zhang[1,3,*]

[1]Key Laboratory of Crop Gene Resources and Germplasm Enhancement, Ministry of Agriculture and Rural Affairs/Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[2]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

[3]Nanjing Agricultural University, Nanjing 210095, Jiangsu, China

[4]Agronomy College/National Key Laboratory of Wheat and Maize Crop Science/Collaborative Innovation Center of Henan Grain Crops, Henan Agricultural University, Zhengzhou 450046, China

[5]Henan Institute of Science and Technology, Xinxiang 453003, China

[6]Agriculture Victoria Research, Department of Economic Development, Jobs, Transport and Resources, AgriBio, Bundoora, VIC 3083, Australia

[7]University of the Chinese Academy of Sciences, Beijing 100049, China

[8]These authors contributed equally to this article.

***Correspondence: Jizeng Jia (jiajizeng@caas.cn), Xueyong Zhang (zhangxueyong@caas.cn)**

https://doi.org/10.1016/j.xplc.2023.100556

ABSTRACT

The centromere is the region of a chromosome that directs its separation and plays an important role in cell division and reproduction of organisms. Elucidating the dynamics of centromeres is an alternative strategy for exploring the evolution of wheat. Here, we comprehensively analyzed centromeres from the *de novo*-assembled common wheat cultivar Aikang58 (AK58), Chinese Spring (CS), and all sequenced diploid and tetraploid ancestors by chromatin immunoprecipitation sequencing, whole-genome bisulfite sequencing, RNA sequencing, assay for transposase-accessible chromatin using sequencing, and comparative genomics. We found that centromere-associated sequences were concentrated during tetraploidization and hexaploidization. Centromeric repeats of wheat (*CRWs*) have undergone expansion during wheat evolution, with strong interweaving between the A and B subgenomes post tetraploidization. We found that CENH3 prefers to bind with younger *CRWs*, as directly supported by immunocolocalization on two chromosomes (1A and 2A) of wild emmer wheat with dicentromeric regions, only one of which bound with CENH3. In a comparison of AK58 with CS, obvious centromere repositioning was detected on chromosomes 1B, 3D, and 4D. The active centromeres showed a unique combination of lower CG but higher CHH and CHG methylation levels. We also found that centromeric chromatin was more open than pericentromeric chromatin, with higher levels of gene expression but lower gene density. Frequent introgression between tetraploid and hexaploid wheat also had a strong influence on centromere position on the same chromosome. This study also showed that active wheat centromeres were genetically and epigenetically determined.

Key words: wheat evolution, centromere repositioning, epigenetic modifications

Zhao J., Xie Y., Kong C., Lu Z., Jia H., Ma Z., Zhang Y., Cui D., Ru Z., Wang Y., Appels R., Jia J., and Zhang X. (2023). Centromere repositioning and shifts in wheat evolution. Plant Comm. **4**, 100556.

INTRODUCTION

The centromere is a constitutive structure of the chromosomes of eukaryotic species; it is required for faithful separation of chromosomes and chromatids during meiosis and mitosis because it ensures connections with spindle fibers (Henikoff et al., 2001). These processes keep the number of chromosomes stable from cell to cell and from generation to generation. For nucleosomes in centromeric regions of most species, the

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

histone H3 protein is partially replaced by CENH3, and the appearance of CENH3 is now regarded as a defining marker of active centromeres. In most plants, including rice (Cheng et al., 2002), maize (Zhong et al., 2002), and sorghum (Jiang et al., 1996; Miller et al., 1998), centromeres are dominated by centromere-specific repeats and retrotransposons. These repeats and retrotransposons are highly variable among species, even among chromosomes within a cell, in terms of both repeat number and type (Lee et al., 2005; Wolfgruber et al., 2009).

Genome sequence and metaomics data have promoted research on centromeres, revealing centromere expansion of maize in the background of oat (Wang et al., 2014), neocentromere formation because of loss of *CentC* in maize (Schneider et al., 2016), and frequent centromere repositioning in Arabideae (Mandakova et al., 2020). Studies have also been performed on inherent epigenetic modifications surrounding centromeric chromatin consisting of blocks of CENH3-containing nucleosomes interspersed with blocks of H3-containing nucleosomes. For example, hypomethylation of DNA in centromeric chromatin compared with the flanking pericentromeric region was revealed in *Arabidopsis* and was correlated with a significantly reduced level of dimethylation of histone H3 at Lys9 (Zhang et al., 2008; Naish et al., 2021). By contrast, DNA sequences associated with functional centromeres can be either hypomethylated or hypermethylated in rice (Yan et al., 2010). In addition, the overlap between dimethylation of histone H3 at Lys4 and centromere protein A (CENP-A) or centromere identifier (CID) regions in humans and *Drosophila melanogaster* (Sullivan and Karpen, 2004), the active genes in centromeric chromatin of rice (Nagaki et al., 2004; Yan et al., 2005), and the slight increase in euchromatic histone modifications such as H3K4me3 in centromeres relative to pericentromeres of *Arabidopsis* (Naish et al., 2021) show that centromeric chromatin is distinct from euchromatin and flanking heterochromatin.

Common wheat (*Triticum aestivum* L.) is a hexaploid species with a complex genome formed from two rounds of natural hybridization between progenitor species. Chromatin immunoprecipitation (ChIP) and immunostaining have established that centromeric repeat of wheat (*CRW*) and *Quinta* are the predominant centromere retrotransposon types in wheat (Liu et al., 2008; Li et al., 2013). A tandem repeat, *Tail*, is also scattered in many wheat centromeres (Kishii et al., 2001). Centromere studies are hindered by sequence complexity, especially for wheat, which has more than 85% repetitive DNA. However, wheat genomes have been deciphered with the development of sequencing and assembly techniques, laying a foundation for exploring the evolution of wheat centromeres. For instance, a shift of approximately 25 Mb was discovered on chromosome 4D in Chinese Spring (Walkowiak et al., 2020). Structural rearrangements and satellite sequence variations have occurred in response to polyploidization in the wheat centromeric region (Su et al., 2019). Wheat centromeric nucleosomes are more sensitive to light than heavy micrococcal nuclease (MNase) digestion conditions (Jordan et al., 2020).

There remains a need for systematic assessment of centromere dynamics during wheat evolution. In our study, we inspected centromere evolution in common wheat and its diploid and tetraploid progenitors using *CRW* and ChIP-seq with a CENH3 antibody. We also assessed the variation in centromeres between two cultivars, Chinese Spring (CS) and Aikang 58 (AK58), in terms of genome sequence and epigenomes; the latter cultivar is widely cultivated in China and has been sequenced and assembled *de novo* (unpublished data). We found mobility of *CRW* throughout various polyploid species in the *Triticum* tribe, a binding preference of CENH3 for younger *CRWs*, an association between methylation and repositioned centromeres, and highly expressed genes in centromeres with more open chromatin. These results update the view that centromeres are fossil genomic regions because they rarely participate in recombination of chromosomes in plant evolution.

RESULTS

Centromere-associated sequences became concentrated during wheat polyploidization

To systematically analyze the dynamics of *CRWs* and *Quintas* in the polyploidization process, we profiled their distribution densities in *Triticum urartu* (Ling et al., 2018), *Aegilops tauschii* (Luo et al., 2017), wild emmer wheat (WEW; *Triticum turgidum* ssp. *dicoccoides*) (Avni et al., 2017), durum wheat (DW; *T. turgidum* ssp. *durum*) (Maccaferri et al., 2019), and common wheat (CW; *T. aestivum*) (International Wheat Genome Sequencing Consortium, 2018). After aligning intact long terminal repeats (LTRs) and the full internal region of *CRWs* and *Quintas* against the genome sequence of wheat and its relatives using BLASTN (Altschul et al., 1990) and rigorous filtering based on over 80% identity and coverage, we found that *CRWs* and *Quintas* became increasingly concentrated, from a relatively round peak to a sharp peak, during polyploidization (Figure 1A). For instance, the highest proportion of *CRWs* and *Quintas* was approximately 0.25 in *T. urartu*, but this value exceeded 0.5 within 50-kb windows in WEW, DW, and CW (Figure 1A). Even so, the content of *CRWs* and *Quintas* was still underestimated, because the content increased from 0.5 in *T. urartu* to nearly 1 in WEW, DW, and CW when only 80% identity was considered (Supplemental Figure 1).

Cytological immune colocalization of CENH3 and centromere-associated DNA sequences was used to further verify *CRW* and *Quinta* dynamics, given the difficulty in genome assembly caused by the highly repetitive sequences. We observed a concentration of *CRW* homologs from diploids to tetraploids or hexaploids in subgenomes A and D, which was consistent with the density profile (Figure 1B). Interestingly, *CRWs* had an extremely faint outline in *Aegilops speltoides*, which diverged from an ancestor of the donor of the CW B subgenome (Li et al., 2022), but there was no significant difference between A and B subgenome chromosomes in tetraploid and hexaploid wheat (Figure 1B). Moreover, *Tail*, the centromeric sequence in *Ae. speltoides*, was dramatically reduced and marginalized in polyploids (Figure 1C). Therefore, we infer that *CRWs* and *Quintas* became the core DNA sequences of centromeres in polyploid wheat.

CRWs proliferated rapidly to unify the three subgenomes in CW

To gain insights into the evolutionary history of allohexaploid wheat, 1552 intact *CRWs*, including 632, 438, and 387 from the

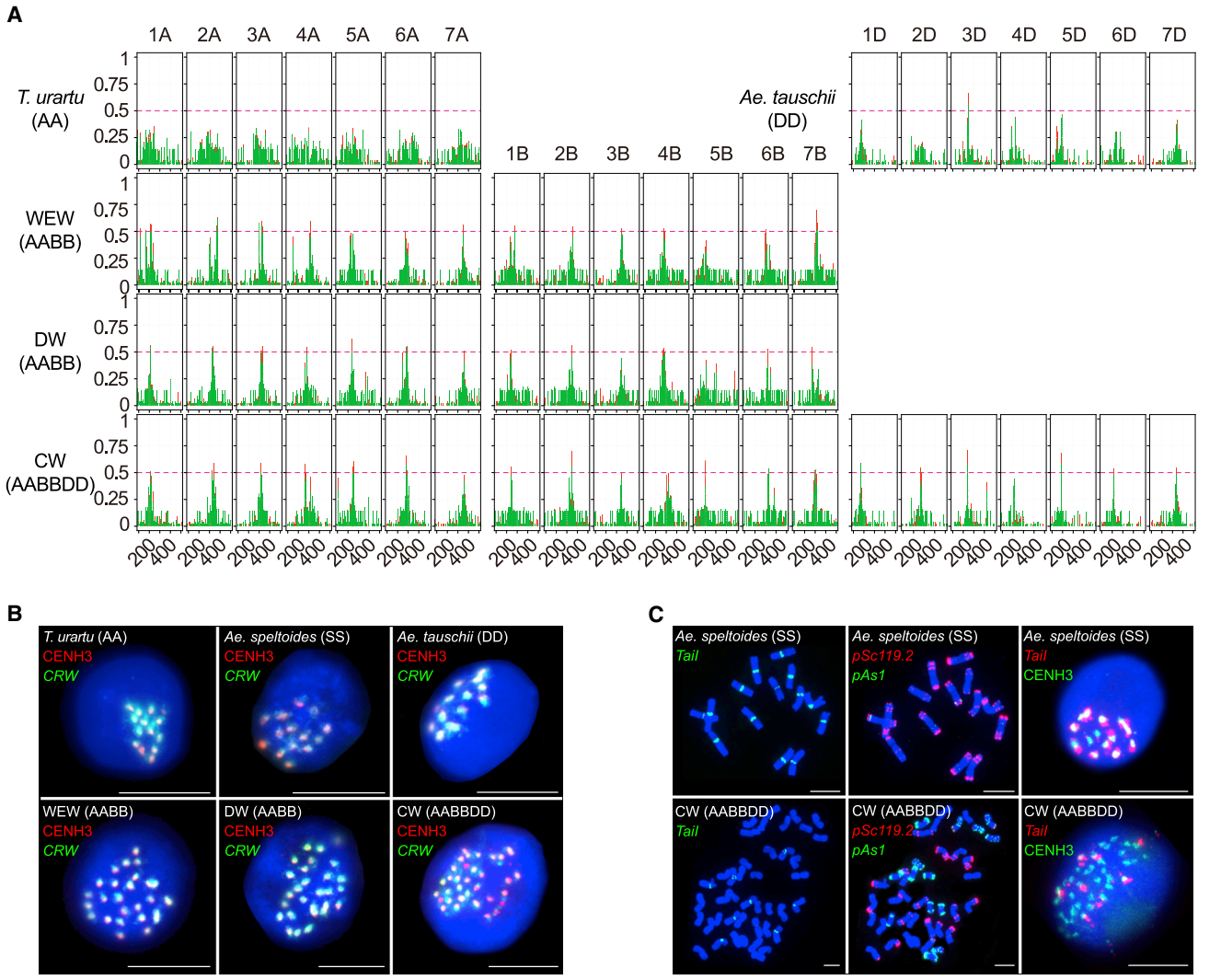

Figure 1. Concentrations of *CRWs* and *Quintas* accompanied by marginalization of *Tail* in centromere during wheat evolution.
(A) Density of *CRWs* and *Quintas* in chromosomes of sequenced wheat and relatives, including *T. urartu*, *T. turgidum* ssp. *dicoccoides* (WEW), *T. turgidum* ssp. *durum* (DW), *T. aestivum* (CW), and *Ae. tauschii*. The y axis indicates the density of LTRs and internal domains of *CRWs* and *Quintas* per 50-kb window. *CRWs* and *Quintas* are shown in green and red, respectively.
(B) Immunocolocalization of CENH3 and *CRWs* in five sequenced species shown in **(A)**, as well as *Ae. speltoides*, the possible ancestor of the B subgenome. *CRWs* and CENH3 are shown in green and red, respectively, after costaining with DAPI. The nucleus is indicated in blue. Scale bars, 10 μm.
(C) Conspicuous marginalization and elimination of *Tail* from diploids to hexaploids. The first and second columns show the signal of *Tail* as well as *pSc119.2* and *pAs1* in the same cell by multiple staining. The last column shows the analysis of colocalization of CENH3 and *CRWs* in the nucleus. The colors are indicated in the images. Scale bars, 10 μm.

A, B, and D subgenomes and 95 from unanchored scaffolds, were extracted from the CS reference sequence v1.0 to construct a phylogenetic tree. Overall, the tree could be roughly divided into three parts based on the A, B, and D subgenomes, and the A and D subgenomes had a closer genetic relationship (Figure 2A). Notably, a fraction of *CRWs* frequently intertwined, resulting in strong connections among younger *CRWs* from the A and B subgenomes in the neighbor-joining tree (Figure 2A).

A systematic phylogenetic tree based on 4687 full-length *CRWs* from wheat and its relatives indicated that intertwining was due to their amplification after tetraploidization (Figure 2B). The reason was that the interwoven node of the A and B subgenomes contains *CRWs* from polyploids but not diploid ancestors. However, the D subgenome remained pure, with no obvious infiltration from the A or B lineage, probably owing to the short time over which it integrated into the AABB genome (Figure 2B). Estimation of insertion times for all *CRWs* also showed that tetraploid and hexaploid wheat had more young *CRWs* (Figure 2C). This was also supported by the younger insertion time of *CRWs* in the D subgenome of CS than in that of *Ae. tauschii* (Figure 2D). Nevertheless, we could not rule out the amplification of *CRWs* in *T. urartu* because it had a distinct peak of expansion of less than 0.5 million years (Figure 2C).

To further explore whether *CRWs* expanded independently or simultaneously, we calculated genetic distances for 294 intact

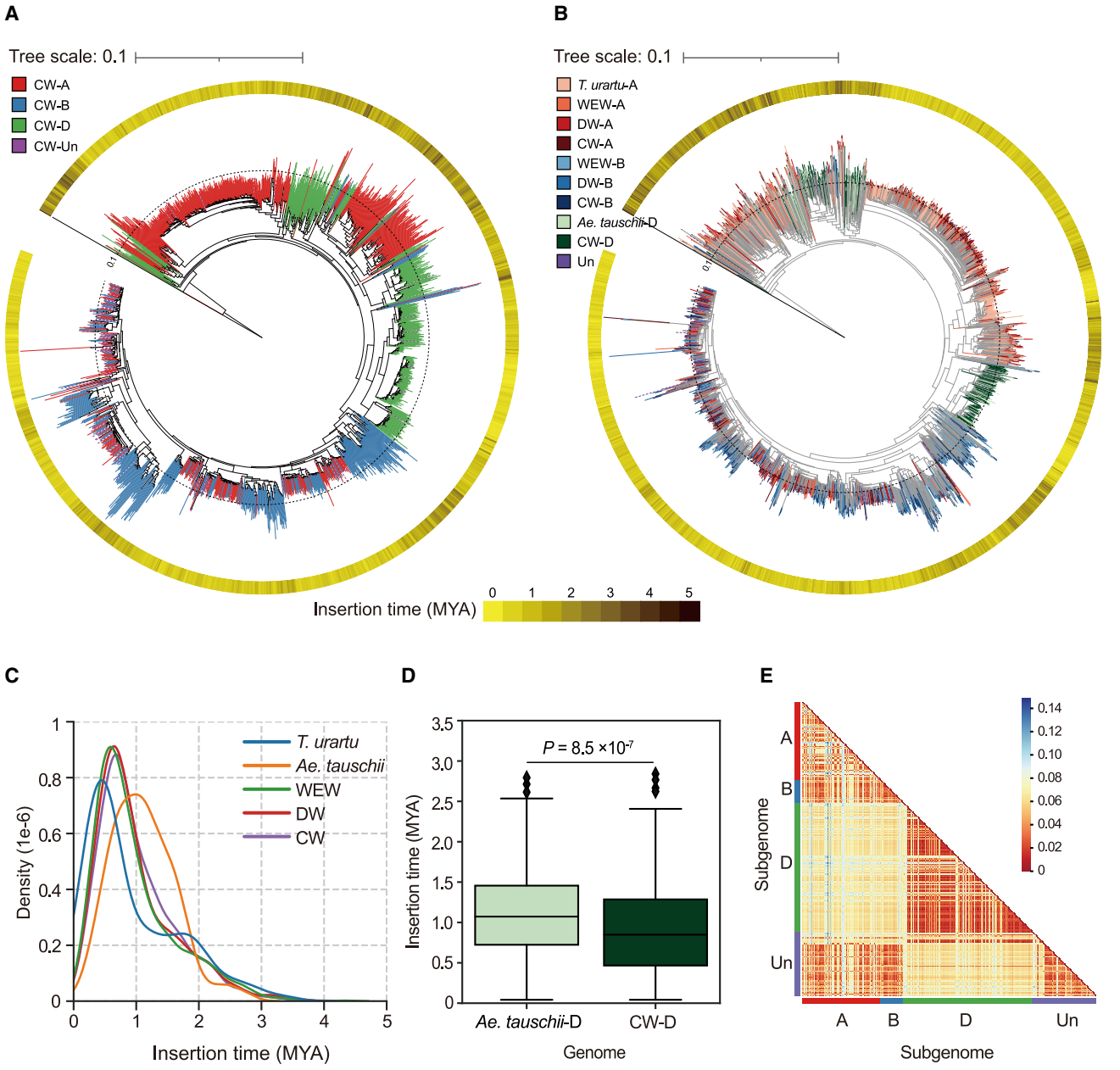

Figure 2. Bursts and interweaving of *CRWs* during wheat polyploidization.

(A) *CRWs* interwoven between the A and B subgenomes as revealed by a phylogenetic tree based on full-length *CRWs* in common wheat (CW). The A, B, and D subgenomes are represented by red, blue, and green. The *CRWs* of unanchored chromosomes are indicated by the purple dashed line. The outermost heatmap represents the insertion time, which is calculated from the divergence of LTRs of *CRWs*.

(B) Interweaving between subgenomes A and B in tetraploids and hexaploids, as presented in a phylogenetic tree of full-length *CRWs* in *T. urartu*, WEW, DW, CW, and *Ae. tauschii*. The different gradients of red, blue, and green (representing the A, B, and D subgenomes) gradually darken with the transition from diploidy to tetraploidy to hexaploidy.

(C) Polyploidization accelerates the expansion of *CRWs* with unique densities in tetraploids and hexaploids compared with diploids in the kernel density estimate plot. However, the existence of young *CRWs* in the genome of the older species *T. urartu* highlights the dynamic nature of retrotransposons in its centromeres. The x axis indicates the insertion time, and the y axis shows the density.

(D) More younger *CRWs* were detected in the D subgenome of CW than in *Ae. tauschii*. This finding further demonstrated that a *CRW* burst occurred in CW. The center values correspond to the medians of the distributions. The *P* value was determined by a nonparametric Mann–Whitney *U* test.

(E) The genetic distance of *CRWs* in the functional centromere revealed that the D subgenome expanded independently, whereas the A and B subgenomes expanded together in CW.

CRWs embedded in CS centromeres. A heatmap showed the close genetic relationship of *CRWs* from the A and B subgenomes, which were distinct from those of the D subgenome, implying that the expansion of *CRWs* in the centromeric region of the D subgenome was independent of the A and B subgenomes (Figure 2E). In summary, these results indicate that

CRWs proliferate during polyploidization, which may facilitate centromere clustering to maintain stable transmission of chromosomes. Such proliferation could also be found in *Quintas* and was younger than that in *CRWs* (Supplemental Figure 2).

CENH3 preferred to bind with chromatin regions enriched with younger *CRWs*

Binding of only some of the *CRWs* to CENH3 was demonstrated by immunocolocalization (Figure 1B). When we performed fluorescence *in situ* hybridization (FISH) with *CRWs*, we found two distinct signals on chromosomes 1A and 2A in WEW, but only the large domain was active for centromere function, showing colocalization with CENH3 (Figure 3A and 3B and Supplemental Figure 3). We further pinpointed the active centromere on chromosome 1A (212.1–215.7 Mb and 217.0–218.7 Mb) embedded in the *CRW*-enriched region of 210.8–220.3 Mb, but the inactive fragment was located at 171.3–177.7 Mb based on the distribution characteristics of *CRWs* and CENH3 (Figure 3C). Similarly, on chromosome 2A, the active centromere (367.0–371.3 Mb) was mapped within 366.2–372.3 Mb, and inactive centromeres were scattered in the wild emmer genome (Figure 3D). Intriguingly, we found that *CRWs* within active centromeres were predominantly younger than those in inactive centromeres of these two chromosomes ($P < 0.01$, Mann–Whitney *U* test) (Figure 3E).

There was no significant difference in gene density or transcription between active and inactive centromeres dominated by *CRWs* on chromosomes 1A and 2A of WEW (Figure 3C and 3D; Supplemental Tables 1 and 2). However, gene densities and transcription levels were lower in the activated centromere regions than in nearby regions. Moreover, the adjacent territories of active centromeres usually had highly expressed genes that seemed to hinder CENH3 deposition (Figure 3C and 3D). We inferred that the lack of genes and transcription provided a favorable chromatin environment for neocentromere activation. Although we could not detect inactive centromeres on metaphase chromosomes in DW or CW by FISH using *CRWs* as probes, nonfunctional *CRW* islands were detected by alignment, which showed perfect genomic synteny but with obvious inversions between species (Figure 3C and 3D). Therefore, we should use caution in subsequent FISH analysis because of the low resolution. By contrast, this active domain region has poor synteny, which is consistent with the energetic feature of new *CRWs* (Figure 3C and 3D). Hence, this finding indirectly supports the view that CENH3 prefers to bind with younger *CRWs*.

Repositioning and shifts of active centromeres on chromosomes 3D, 4D, and 1B in AK58 compared with CS

Physical mismatches between the CENH3 binding regions precipitated from AK58 and CS on chromosomes 3D, 4D, and 1B strongly suggested physical shifts of active centromere regions in these two cultivars (Figure 4). Here, we defined the *CRW*-containing active site as the ancestral centromere and the site with fewer or no *CRWs* as the new centromere, as *CRWs* colocalized with CENH3 in *T. urartu* and *Ae. tauschii* (Figure 1B). The most extreme active centromere repositioning was detected on chromosomes 3D and 4D, where the active centromeres in AK58 and CS defined by CENH3 ChIP-seq reads were mapped to a quite different genomic region without any synteny (Figure 4A and 4B). On chromosome 3D, the centromere of AK58 (238.2–245.5 Mb) had undergone an apparent shift of approximately 8 Mb compared with the ancestral centromere (246.4–255.0 Mb) defined by ChIP-seq reads precipitated from CS. Furthermore, the active centromere of AK58 contained few *CRWs*, even though there were a large number of *CRWs* (248.3–255.5 Mb) without CENH3 binding (Figure 4A). Similarly, an active centromere shift of approximately 22 Mb was detected on chromosome 4D of CS, where CENH3-binding reads from CS were mapped to 182.5–188.6 Mb, whereas those from AK58 were mapped to 204.4–210.5 Mb (Figure 4B). Interestingly, no obvious *CRWs* were detected in the active centromere region of chromosome 4D in CS, which was consistent with previous reports (Koo et al., 2015; Walkowiak et al., 2020).

The most remarkable shift was found on the 1RS/1BL Robertsonian translocation chromosome in AK58, where the active centromere region has moved to the 1RS, basically abandoning the 1BL centromere domain (Figure 4C). Colocalization of the retrotransposon *Bilby* from rye and CENH3 on this 1RS/1BL chromosome further demonstrated that its centromeric DNA is mainly composed of *Bilby* (Supplemental Figure 4), consistent with recently reported results (Karimi-Ashtiyani et al., 2021). We also observed trends of CENH3 binding domains shifting to gene-poor regions on chromosomes 1RS/1BL, 3D, and 4D (Figure 4 and Supplemental Figure 5). This might be related to some recruitment mechanism that promotes the expression of genes associated with centromeres during breeding.

Active centromeres usually have higher CHH and CHG methylation levels

Notably, some "islands" of *CRWs* have lost their ability to bind CENH3 (see such an island on chromosome 3D of AK58 and on chromosome 4D of CS) (Figure 4A and 4B). The reason for this centromere repositioning remains to be explored. We suspected a shift probably related to the degeneration of *CRWs* or changes in centromeric chromatin characteristics. First, we constructed a clustering tree based on *CRWs* extracted from *CRW*-enriched active centromeric regions and from inactive centromeric regions with synteny. On chromosome 3D, *CRWs* of the active centromere (236.9–243.6 Mb in CS) and inactive centromere (246.4–255.0 Mb in AK58) were not distinguished from each other on the basis of homology (Supplemental Figure 6A). Similarly, the *CRWs* of active centromeres in AK58 (208.6–214.8 Mb) and the syntenic region in CS (204.4–210.5 Mb) on chromosome 4D also did not differ (Supplemental Figure 6B). Therefore, we suggest that the degradation of *CRWs* is not the reason for establishment of *de novo* centromeres in *CRW*-depleted regions.

The activity of transposons is negatively correlated with DNA methylation. Using bisulfite sequencing, we detected hypermethylation of CHG and CHH on the DNA sequence in the active centromeric region in CS or AK58 on chromosome 3D, whereas CG was hypomethylated (Figure 5A and Supplemental Figures 7 and 8). On chromosome 4D, the active centromere

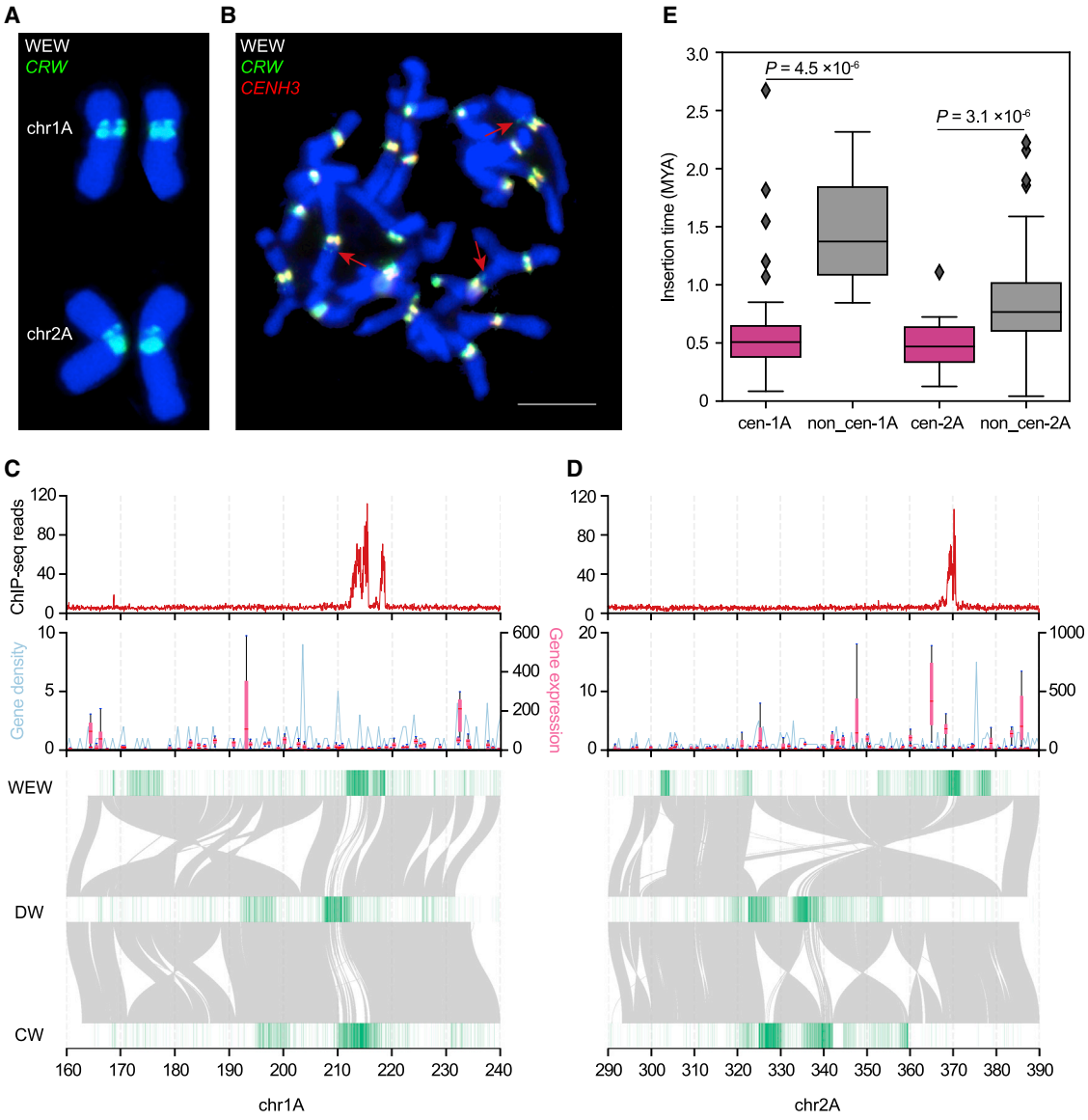

Figure 3. Active centromeres contain more younger *CRWs* than inactivated centromeres in the dicentric chromosomes 1A and 2A in WEW.

(A and B) Two *CRW*-rich regions were detected on chromosomes 1A and 2A in WEW, but only one *CRW*-rich region bound CENH3 in each of the dicentric chromosomes. *CRWs* and CENH3 are marked by green and red, respectively. Scale bar, 10 μm.

(C and D) Remarkable adjustment of centromeres to genomic regions containing sparse and low-expressed genes on chromosomes 1A and 2A during evolution from WEW to CW. Top: each centromere was identified using the number of ChIP-seq reads in 50-kb windows per million reads with CENH3 antibody precipitation. Center: the gene density at the centromeric region was calculated as the number of high-confidence genes in 50-kb windows. The expression (TPM) of genes was calculated using RNA-seq data from 20 different combinations of WEW tissues collected at various developmental stages. Bottom: extraordinary adjustment of centromeres is indicated by DNA sequence synteny among WEW, DW, and CW; the green lines represent *CRWs*.

(E) Preference of CENH3 to bind younger *CRWs*. The center values correspond to the medians of the distributions. The *P* value was determined by a nonparametric Mann–Whitney *U* test.

regions of CS were much more extensively methylated in the CHG and CHH contexts than the inactive centromere regions defined by CENH3-binding sites from AK58, and a complementary methylation pattern was shown in corresponding syntenic regions in AK58 (Figure 5A and Supplemental Figures 7 and 8). Similarly, on chromosomes 6A and 7B of WEW and DW, where the centromeres had shifted, we found that the active centromere of WEW had high CHG and CHH methylation and low CG methylation (Figure 5B and Supplemental Figures 9 and 10). However, these sequences did not bind to CENH3 in DW, exhibiting a shift in which CHH and CHG methylation was reduced but CG methylation was elevated (Figure 5B and Supplemental Figures 9 and 10). Thus, methylation and demethylation could be driving forces for centromere repositioning.

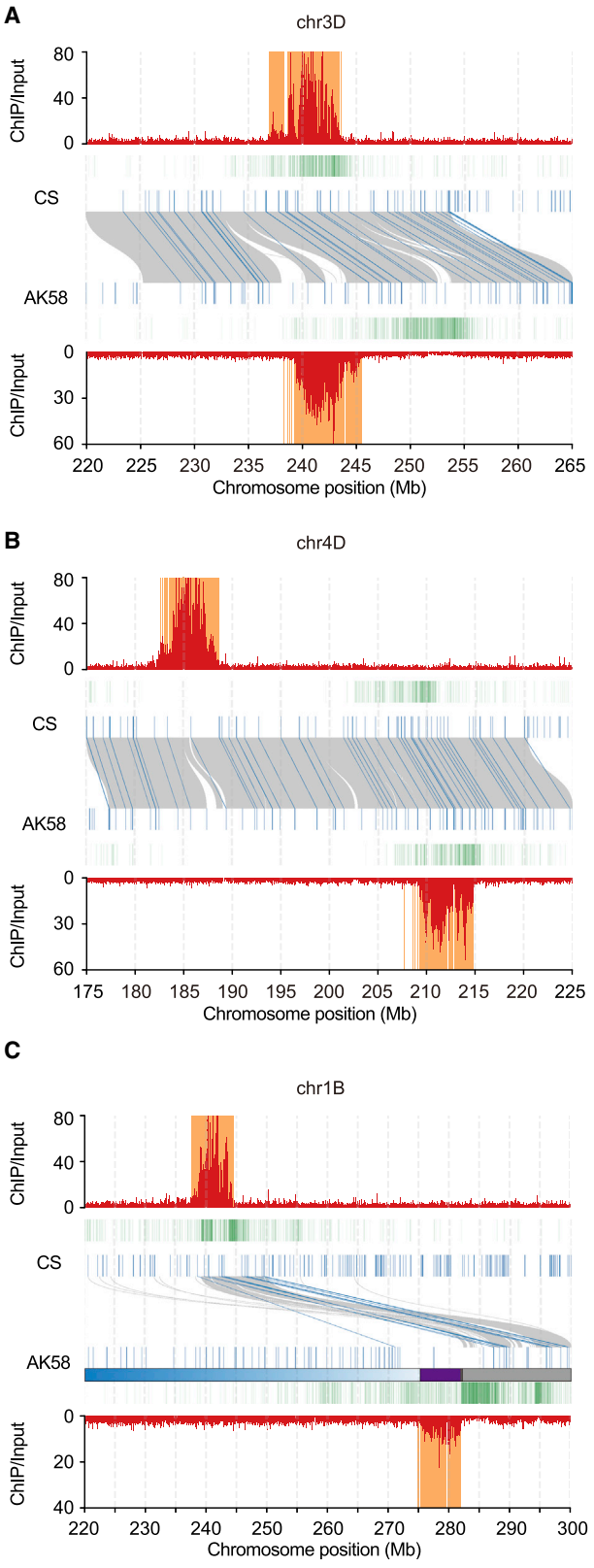

Figure 4. Shifts of active centromeres in chromosomes 3D, 4D, and 1B were detected between CS and AK58.

(A) The centromere of chromosome 3D shifts to a *CRW*-poor island in AK58 compared with CS.

(B) Repositioning of the centromere to a region without *CRWs* on chromosome 4D in CS that is approximately 22 Mb away from the active centromere region of AK58.

(C) The active centromere in the 1RS/1BL Robertsonian translocation chromosome shifts to 1RS in AK58. The 1RS part, 1BL part, and centromeres are represented by blue, gray, and purple bars.

The gold blocks indicate active centromere subdomain binding with CENH3. The normalized read count ratio between ChIP-seq and input in 1-kb windows is shown in the red histogram (CPM normalization). The homologous genes and syntenic regions between CS and AK58 are anchored by blue lines and gray curves, respectively. Green lines represent *CRWs*.

Genes colocalized with H3 in centromeric regions were expressed at higher levels because of the greater openness of chromatin

We used ATAC-seq to investigate chromatin accessibility around the centromeric region in AK58, which was identified by ChIP with an CENH3 antibody. In most cases, the read coverage of centromeric regions was higher than that of pericentromeric regions, especially for the D subgenome chromosomes (Figure 6A). For example, on chromosome 3D, the centromere of AK58 showed a higher degree of chromatin openness, whereas the homologous region in CS was inactive and exhibited reduced chromatin accessibility. Open chromatin was also revealed on centromeres of chromosome 4D in CS and AK58, showing a contrasting profile in the collinear region (Supplemental Figure 11). All chromosomes, including chromosome 6B of CS, which had lower CHG methylation in the centromeric region (Supplemental Figure 7), showed a higher degree of chromatin openness in this region than in the surrounding regions (Supplemental Figures 11 and 12).

We next used two sets of RNA-seq data from leaf and floret meristem tissues to investigate the expression levels of 115 genes located in centromeric regions by comparison with 115 genes randomly selected within 10 Mb on both sides of the flanking region. We found that the median gene expression level in the centromeric region was higher than that in the flanking region, but no significant difference was shown (Figure 6B and 6C). Interestingly, when we divided the genes into two categories according to whether they were located in the H3 subdomain or the CENH3 subdomain of centromeric regions, we found that most of the 33 genes located in the CENH3 subdomain were not expressed, and the average expression levels in differentiated leaf and floret tissues were 2.373 and 4.987, respectively (Figure 6B and 6C). However, compared with those in the CENH3 subdomain, the 82 genes embedded in the H3 subdomain had higher expression levels, with an average of 21.683 in floret tissues and 9.665 in leaves (Figure 6B and 6C). Moreover, the accessibility in the TSS of genes covered by H3 nucleosomes was higher than that of CENH3 in the centromeric region, and histone modification characteristics were also present in actively expressed genes (Supplemental Figure 13; Supplemental Table 3).

To investigate whether centromere relocation has an effect on gene expression, differential expression analysis was performed with CS as a control. Among 83 genes involved in centromere relocation, only 4 were differentially expressed (Supplemental Table 4). Among them, TraesAK58CH1B01G123100 (lactose glutathione lyase) was upregulated and embedded in the H3 subdomain of

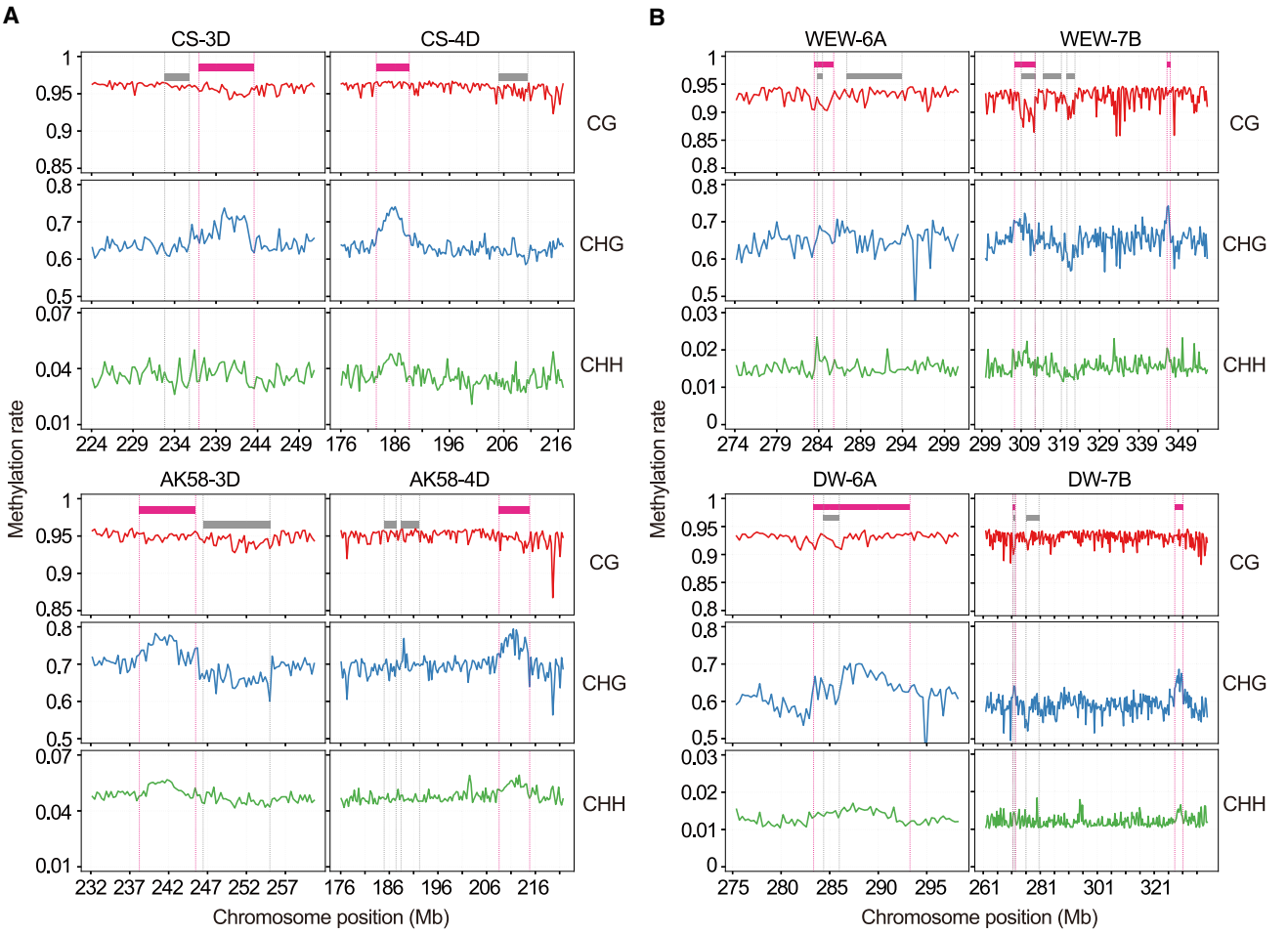

Figure 5. Active centromere regions are usually associated with higher CHH and CHG methylation.
(A) Complementary distribution of methylation among AK58 and CS indicates that CHG methylation may contribute to centromere movement in chromosomes 3D and 4D.
(B) A high CHG methylation status was observed in WEW and DW at active centromeres, and this status was reduced in the syntenic region in chromosomes 6A and 7B.
The active centromere region is marked by bars and dotted lines in pink, whereas the inactive centromere region is marked by bars and dotted lines in gray. The methylation rate was calculated as the average ratio of the number of methylations to all detections in 300-kb windows.

the centromeric region in AK58, whereas homologous genes were not found in the centromeric region in CS. TraesAK58CH1B01G123900 (Ras-related protein) and TraesAK58CH3B01G242000 expression was downregulated from the H3 subdomain of the centromeric region in CS to the noncentromeric region in AK58. Even so, the expression of TraesAK58CH6B01G255200 was downregulated after switching from the noncentromeric region in CS to the H3 domain of the centromeric region in AK58. These results indicate that there are expressed genes in the centromeric region, especially in the H3 subdomain, which is also consistent with greater chromatin openness.

DISCUSSION

Active centromeres have repositioned and shifted frequently during wheat evolution, and the mechanism is poorly understood

Observations of centromere repositioning in plants have increasingly been reported, although not as frequently as in primates (Montefalcone et al., 1999; Lo et al., 2001; Ventura et al., 2001, 2004, 2007) and Equidae (Carbone et al., 2006). For example, changes in centromere positions of chromosomes 6 and 7 have been revealed between cucumber (*Cucumis sativus* L.) and melon (*Cucumis melo* L.) by comparative FISH mapping (Han et al., 2009). The functional centromere of maize chromosome 3 was shifted to gene-poor regions occupying a larger area in the oat genome background (Wang et al., 2014). A distinct CENH3-binding domain on chromosome 3 was detected in five different maize lines (Zhao et al., 2017). Strong human selection for centromere-linked genes drove rapid centromere repositioning in maize (Schneider et al., 2016). Centromeric DNA sequence deletion led to *de novo* centromere formation in the pericentromeric region of rice chromosome 8 (Xue et al., 2022).

Findings related to wheat centromeres have been very clear in terms of sequence composition, with the centromeres being composed mainly of two retrotransposon types, *CRWs* and *Quintas*, but there are relatively few studies on centromere

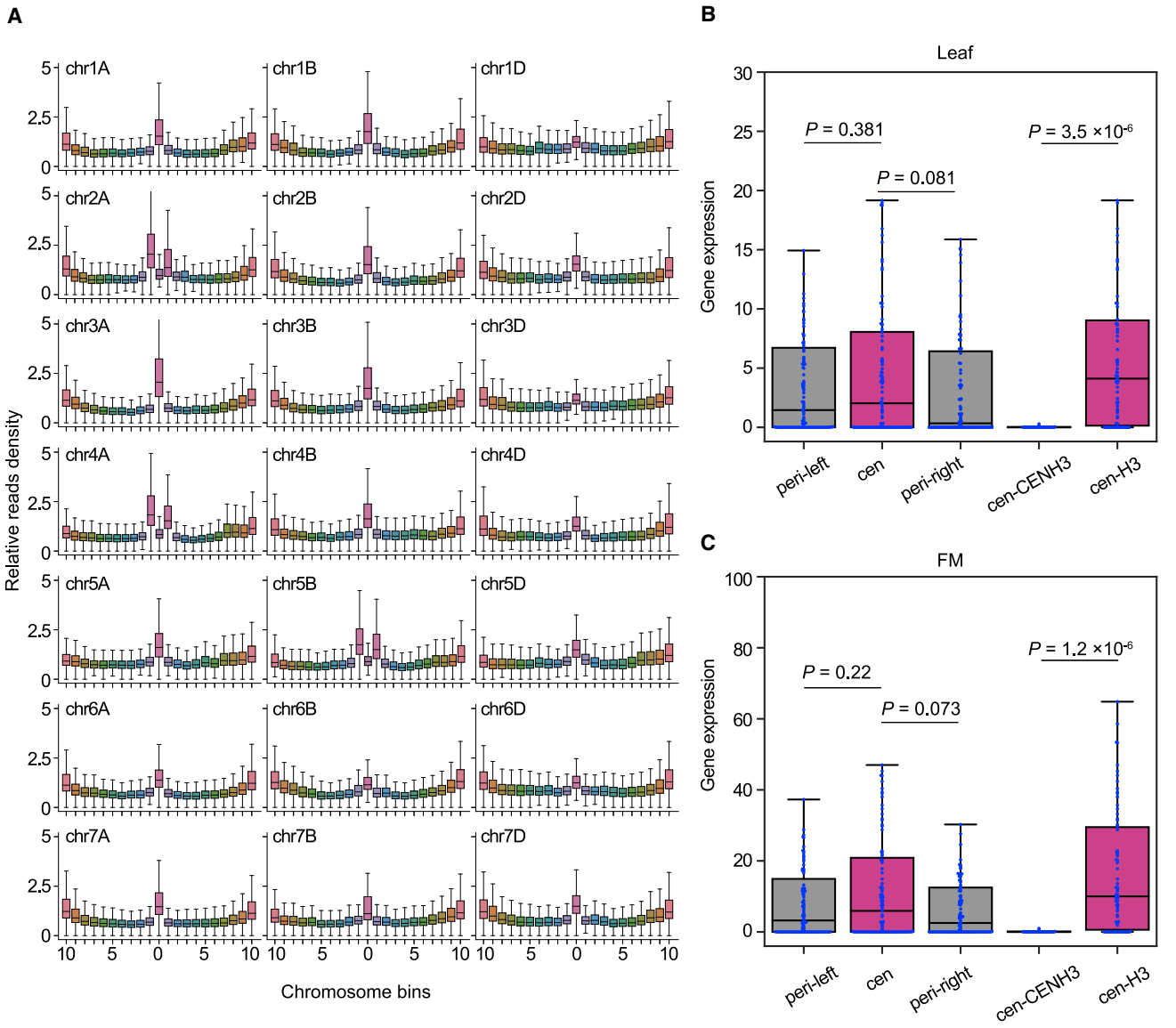

Figure 6. Highly expressed genes in centromeric regions are embedded in relaxed chromatin regions.

(A) The chromatin accessibilities of centromeres are higher than those of nearby regions. Each chromosome except the centromere is split into 20 bins. The ATAC-seq reads per 10-kb window in each bin were calculated.

(B and C) Genes located within histone 3 were usually expressed at higher levels than those located within CENH3 in centromeric regions in the leaf **(B)** or floret meristem **(C)**. Gene expression was normalized by the trimmed mean of M values method. The *P* value was determined by a nonparametric Mann–Whitney *U* test.

evolution at the genome level. The phenomenon of centromere movement on chromosome 4D was reported in wheat, but no clear reason was given (Walkowiak et al., 2020). In this study, we found that active centromeres have more younger *CRWs* than inactive centromeric regions, such as on chromosomes 1A and 2A in WEW. Furthermore, structural variations were obvious between the active and inactive centromeres, and frequent inversions associated with wheat polyploidization and domestication were detected in centromeric regions on these two chromosomes (Figure 3C and 3D).

We found centromere movement events on six chromosomes: 6A and 7B in WEW and DW and 3D, 4D, 1B, and 6B in CS and AK58. These events can be divided into two categories. In one, the novel active centromere and old inactive centromere contain *CRW* homology sequences, such as those on chromosomes 6A and 7B in WEW and DW and on 6B in CS and AK58 (Figure 7 and Supplemental Figures 14 and 15). Centromeric relocation might have occurred during the domestication of tetraploid wheat, because the new centromere region is rich in *CRWs* and there are obvious sequence deletions and insertions. We speculate that the time required for this process is relatively long. In the other category, no or few *CRW* homology sequences were detected in the novel active centromeres, as observed on chromosomes 3D and 4D in CS and AK58 (Figures 4A, 4B, and 7). However, the novel centromere on chromosome 3D in AK58 had more *CRWs* than 4D in CS, and two large insertions were detected (238–240.3 Mb and 242.1–244 Mb), which

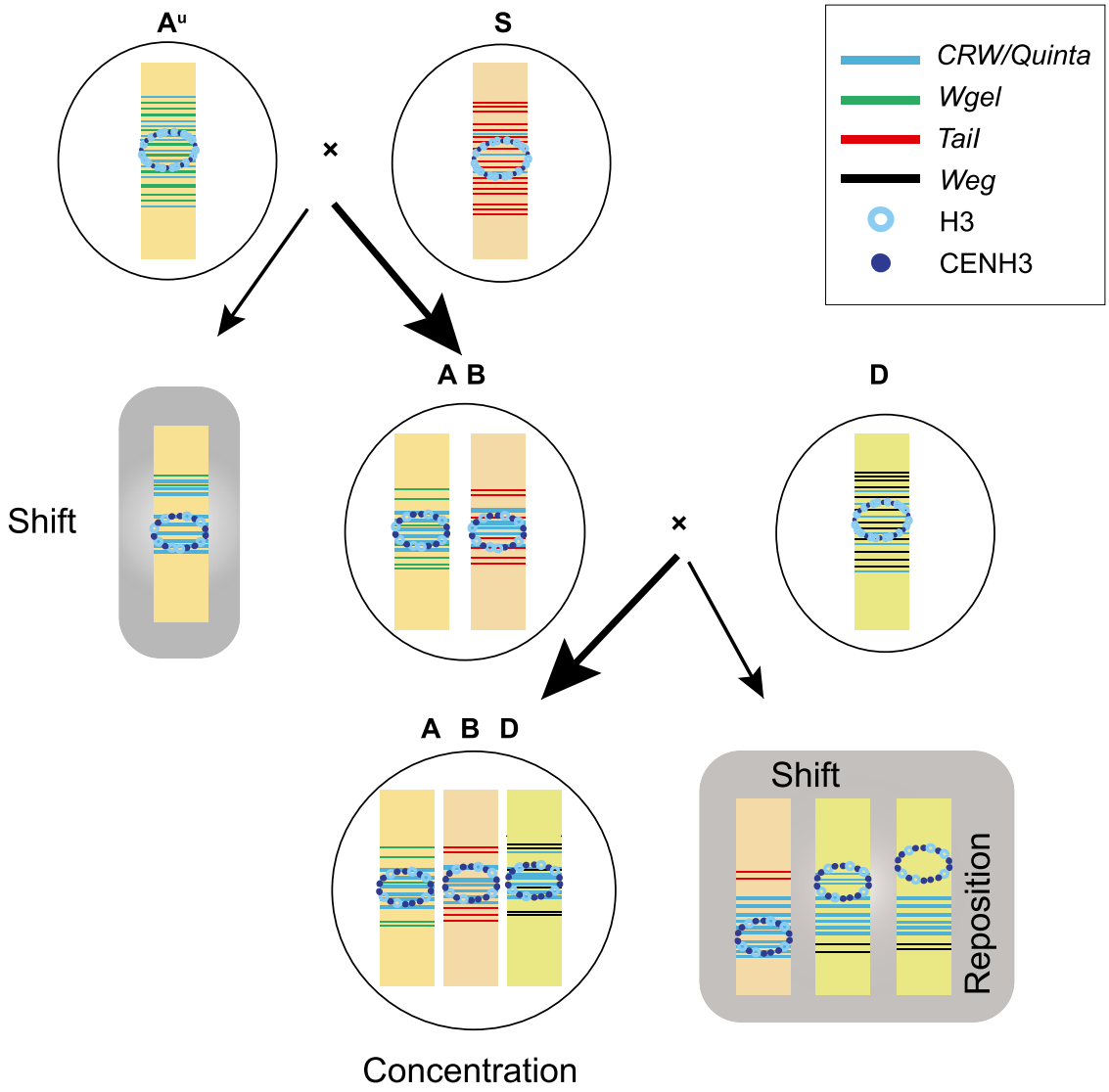

Figure 7. Basic illustrations of centromere concentration, shifts, and repositioning in wheat evolution.
During polyploidization, *CRWs* and *Quintas* expanded and concentrated gradually in centromeres, replacing the A, B, and D subgenome centromeric sequences *Wgel*, *Tail*, and *Weg*. In tetraploid wheat, centromeres shifted to the young *CRWs* on WEW chromosomes 1A and 2A. In hexaploid wheat, centromeres covered by distinct *CRW* regions or *CRW*-depleted regions were found in CS and AK58 on the same chromosome. The thickness of the arrows indicates major and minor types, respectively.

may be related to chromosome rearrangement. Higher CHG methylation was found in the active centromeric regions than in the inactive regions on chromosomes 3D and 4D (Figure 5A). This implies that CHG methylation is not fully related to centromere retrotransposons because there are no obvious *CRW* homology sequences in the active centromeric region. Moreover, there is a high degree of chromatin openness in all active centromeres (Figure 6A and Supplemental Figure 12). The methylation profile in the functional *CRW*-enriched region was distinct from that of the inactive region on chromosomes 1A and 2A of WEW, which also indicated higher CHG methylation in centromeres (Supplemental Figure 16). In a previous analysis, centromere loss and relocation were explained by chromosome rearrangements or centromere inactivation (Lysak, 2014). Here, we propose that hypomethylation of the CG context and hypermethylation of the CHG and CHH contexts in the centromeric region and more open chromatin are closely related to the active centromere, although we cannot yet explain the mechanism of centromere movement.

D subgenome centromeres are still under rapid reconstruction in CW

Emmer wheat arose approximately 500 000 years ago through natural hybridization between an *Ae. speltoides* (SS)-type diploid and *T. urartu* (AA) in the Fertile Crescent. It was then domesticated into cultivated *T. turgidum*. Approximately 8000–10 000 years ago, *T. turgidum* hybridized with *Ae. tauschii* to form CW, an excellent model system for studying allopolyploidization and genome evolution. Studies in this field have primarily focused on changes in genes, such as duplication, deletion, pseudofunctionalization, and methylation, and only a few papers have

directly addressed the topic of centromere evolution (Liu et al., 2008; Li et al., 2013; Su et al., 2019). The recent genome assembly of cultivars allows us to explore centromeres more realistically at the genome level with ChIP-seq results. In this study, we delineated the locations of centromeres in *Ae. tauschii*, WEW, DW, CS, and AK58, which, in turn, can aid in correction of genome assemblies and provide markers for gene cloning (Supplemental Tables 5–9; Supplemental Figures 17 and 18). We found that D subgenome chromosome centromeres tend to have fewer *CRWs* and *Quintas* than A and B subgenome centromeres (Figure 1A). In addition, in most chromosomes of the A and B subgenomes, the active centromeric domains are located within regions rich in *CRWs* and *Quintas*; however, in some of the D subgenome chromosomes, such as 3D in AK58 (Supplemental Figure 15), the active centromeres are outside of regions rich in *CRWs* and *Quintas*. This implies that some other repeats or retrotransposons, such as *Weg*, are partially loaded with CENH3 in wheat (Li et al., 2013; Figure 7). Our interpretation is that these repeats are being replaced by *CRWs* and *Quintas*, which is why the physical size of D chromosome centromeres in genome sequences is usually larger than the expected size based on *CRW* or *Quinta* FISH signals. The D subgenome chromosomes usually have weaker and smaller signals for *CRWs* and *Quintas* (Liu et al., 2008; Li et al., 2013). All of these clues indicate that D subgenome centromeres have been engaging in a rearrangement process at a more rapid rate than A and B subgenome centromeres as *CRWs* and *Quintas* are replacing other repeats. We propose that the centromere remains young and variable relative to other regions of the chromosome, accompanied by a large number of retrotransposon insertions and recombination events.

METHODS

Plant materials

Seven samples were used in this study: three diploids, two tetraploids, and two hexaploids. *T. urartu* accession G1812 (AA) and *Ae. tauschii* accession Y2282 (DD) were used as representatives of the progenitors of the A and D subgenomes of hexaploids, respectively. *Ae. speltoides* accession Y2032 (BB), *T. turgidum* accessions Zavitan (AABB) and Svevo (AABB), and *T. aestivum* CS (AABBDD) and AK58 (AABBDD) were also used.

The distribution density of *CRWs* and *Quintas* on chromosomes

To calculate the density of *CRWs* and *Quintas*, we first downloaded the consensus sequences (RLG_Taes_Cereba_consensus-1 and RLG_Taes_Quinta_consensus-1) from the Transposable Elements Platform (http://botserv2.uzh.ch/kelldata/trep-db/TEClassification.php; Wicker et al., 2002). Because transposons are usually nested or truncated, we divided the consensus sequence into two parts to ensure accuracy and sensitivity: the LTR and the internal domain. Next, we aligned the split sequences to the genome using BLASTN (Altschul et al., 1990) and filtered based on the 80-80-80 rule proposed by Wicker et al. (2007). We then calculated the sequence density in 50-kb bins using BEDTools (Quinlan and Hall, 2010); i.e., the mapped length (kb)/50 kb. Finally, a stacked bar plot was created using Matplotlib (Hunter, 2007).

FISH and immunolocalization

Root tips of seedlings cultured on wet filter paper at 23°C were collected when the root length was 1–2 cm. Metaphase chromosome slides for FISH were prepared according to a previously described procedure with some modifications (Kato et al., 2006). The root tip was cut and treated with nitrous oxide for 2 h to induce metaphase, followed by fixation in 90% acetic acid for 8 min. The root sections were digested for 50 min in a 37°C water bath using a mixed enzyme solution (1% pectolyase Y23 and 2% cellulase Onozuka R-10). Finally, the root tip was crushed with a dissecting needle, resuspended in 22 μl of acetic acid, and dropped at 8 μl onto a glass slide. Probes for FISH were prepared by PCR amplification with primers (Supplemental Table 10), followed by labeling via nick translation. Digoxigenin-11-dUTP and biotin-16-dUTP were used to label sequences, and the hybridization signals were detected with anti-digoxigenin-rhodamine (11207750910, Roche) and fluorescein-conjugated goat anti-biotin (SP-3040, Vector Laboratories), respectively. Sequential FISH using *psc119.2* and *pAs1* was performed to distinguish chromosomes. To confirm chromosomes 1A and 2A in WEW, Oligo1AS (an oligo probe specific to 1A) and Oligo428 (a probe that can recognize the 2A chromosome) (Tang et al., 2018) were also used.

Immunocolocalization was performed following a previously described procedure with minor modifications (Zhao et al., 2019). Root tips were fixed in 4% paraformaldehyde in 1× phosphate-buffered saline (PBS) for 20 min and then washed with 1× PBS. The root tip was placed under a coverslip and broken with a dissecting needle, frozen in liquid nitrogen, and dried with 70%, 95%, and 100% ethanol. An anti-rabbit polyclonal antibody against the peptide of the specific N-terminal tail of *Ta*αCENH3, "CARTKHPAVRKTK," was prepared for identification of wheat centromeres. Approximately 1 μg of rabbit anti-CENH3 diluted with 100 μl of Tris/NaCl/blocking reagent buffer (TNB) was added to each slide, and the slide was incubated at 37°C for 3 h and washed with 1× PBS. Approximately 1 μl of fluorescein isothiocyanate anti-rabbit immunoglobulin G (H+L) diluted 1:100 with TNB was added, followed by incubation for 1 h and washing with 1× PBS. To detect the colocalization of CENH3 and the DNA sequence, slides were fixed in 4% paraformaldehyde for 20 min and then dehydrated in 70%, 95%, and 100% ethanol before being used for FISH. Signals were detected under a fluorescence microscope (Axio Imager.Z2, Carl Zeiss, Germany) and captured by a charge-coupled device camera (AxioCam HRM, Carl Zeiss, Germany). Images were processed with Adobe Photoshop CS.

Phylogenetic analysis and insertion time of *CRWs*

We retrieved full-length LTR retrotransposons (LTR-RTs) using EDTA (Ou et al., 2019), which can strictly filter the results identified by LTR-harvest (Ellinghaus et al., 2008) and LTR_finder (Xu and Wang, 2007). Taking into account the sequence variation within a family, we selected the three highest-content LTRs and internal sequences as representatives of *CRW* by a BLASTN search using RLG_Taes_Cereba_consensus-1 as the subject and the LTR-RT cluster sequence as the query. Intact *CRWs* were then retrieved by BLASTN searches of full-length retrotransposons using the representatives as subjects, followed by filtering. The filtering step included the following criteria: (1) identity greater than 80%, (2) coverage greater than 80%, (3) sequence with no more than one internal domain or two LTR sequences, and (4) sequence length variation no longer than a quarter of the intact representative.

The ages of *CRW* insertions were estimated using the formula age = distance/(2 × mutation rate), with a mutation rate of 1.3×10^{-8}, and distance was calculated based on the divergence between the 5′ and 3′ LTRs, applying the JC69 correction. To construct a phylogenetic tree, we used MUSCLE (Edgar, 2004) to perform multiple sequence alignment for all full-length *CRWs* in fast mode (-maxiters 2 -diags). A neighbor-joining tree was built from the output using MUSCLE (-maketree -cluster neighborjoining) and visualized with iTOL (Letunic and Bork, 2021). We estimated the genetic distance of intact *CRWs* using the EMBOSS package distmat (Rice et al., 2000), applying Jin-Nei Gamma correction and visualizing the results with Matplotlib.

ChIP-seq assay

To isolate centromeric DNA sequences, ChIP against CENH3 was performed following an established procedure with slight modifications

(Nagaki et al., 2003). We conducted ChIP-seq with a biological replicate on one hexaploid (AK58), two tetraploids (Zavitan and Svevo), and two diploids (G1812 and Y2282) and combined ChIP-seq data precipitated from CS and other datasets from previous studies (SRR1686799 and SAMN11655702; Guo et al., 2016; Zhao et al., 2019) to study centromere evolution. The precipitated fragments were subsequently sequenced using the Illumina HiSeq X 10 or NovaSeq system to generate 150-bp paired-end reads. Reads generated from ChIP-seq were aligned to the genome to expose the centromere-associated region using the Bowtie 2 alignment program (Langmead and Salzberg, 2012). Only reads with a mapping quality equal to or greater than 30, considered unique sequences, were retained for further analysis. CENH3-binding subdomains were identified with SICER (Zang et al., 2009) using 1-kb windows, with 3-kb gaps allowed in the defined subdomains. The consistency of replicates was shown by the fraction of reads in peaks and a Venn diagram of peaks (Schmitz et al., 2022; Supplemental Table 11; Supplemental Figure 19). CENH3-binding subdomains were integrated as centromere regions based on the benchmark that separation was less than 500 kb.

Centromere comparative analysis

ChIP-seq data from CS and AK58, as well as Zavitan and Svevo, were used to characterize the dynamics of centromeres during domestication and evolution, mainly in relation to centromere shifting. First, to reduce the impact of differences in the degree of assembly between genomes, ChIP-seq data were aligned to the genome of the compared species to define the inactive centromere region (i.e., CENH3-enriched peaks from CS aligned to AK58) as indicated in the previous section. Then, the moving features were determined by the boundaries of the centromere region. Only dramatic shifts were recorded by manual inspection. Finally, the sequences surrounding the centromeric region were aligned using MUMmer (Marcais et al., 2018) to track major differences in evolution, and the best hit for coding sequences between genomes was applied using BLASTN with an E value of 10^{-5} to verify the differences.

Genomic DNA bisulfite sequencing and data analysis

Genomic DNA was extracted from seedling leaves. Bisulfite treatment of genomic DNA was performed using the Zymo EZ DNA Methylation Lightning Kit. With this method, nonmethylated cytosine nucleotides were converted to uracils and read as thymines when sequenced. Methylated cytosines protected from conversion were still read as cytosines. In brief, 50–100 ng of purified genomic DNA was treated with Zymo Lightning Conversion Reagent in a thermal cycler for 8 min at 98°C, followed by 60 min at 54°C. The bisulfite-treated DNA was purified on a spin column and used to prepare the sequencing library using the EpiGnome Kit (Epicentre). In this procedure, bisulfite-treated single-stranded DNA was randomly primed using a polymerase able to read uracil nucleotides to synthesize DNA containing a specific sequence tag. The 3′ ends of the newly synthesized DNA strands were then selectively tagged with a second specific sequence, resulting in di-tagged DNA molecules with known sequence tags at their 5′ and 3′ ends. These tags were then used to add Illumina P7 and P5 adapters by PCR at the 5′ and 3′ ends, respectively, of the original DNA strand. Only the complement to the original bisulfite-treated DNA was used as the sequencing template; thus, the resulting read 1 was always the same sequence as the original bisulfite-treated strands. The bisulfite genome DNA libraries were then sequenced on the Illumina NovaSeq platform to produce 150-bp paired-end reads.

For data analysis, paired-end sequencing reads were aligned to the genome with the default settings of the Bismark program (version 0.19.0), and only the best unique alignments were reported (Krueger and Andrews, 2011). The extent of methylation of each cytosine in all three contexts (CG, CHG, and CHH) was determined using the Bismark methylation extractor implemented in Bismark (version 0.19.0). Only cytosines with a depth of at least three in all libraries were considered. The methylation ratio of each site was calculated as the number of methylated cytosines (mCs) divided by the number of reads covering the position (Li et al., 2019).

For genome-size scale analysis, the chromosome was divided into 300-kb windows, and the average methylation level of each window was calculated using the BEDTools suite (Quinlan and Hall, 2010). Custom Python scripts were used to rearrange data sets and visually represent the methylation level of the centromere region.

RNA-seq

RNA-seq reads of AK58 leaf and floret meristems were analyzed. The sequences were first filtered using fastp software (Chen et al., 2018). Second, the sequences were aligned to the genome using HISAT2 (Kim et al., 2015), and only unique reads were retained. Read counts were calculated for each gene with featureCounts (Liao et al., 2014) and normalized by the trimmed mean of M values method with edgeR (Robinson et al., 2010). Gene expression levels of Zavitan and Svevo normalized by TPM were downloaded from the WheatOmics website (http://202.194.139.32/expression/index.html; Ma et al., 2021).

ATAC-seq

ATAC-seq was performed as described previously (Lu et al., 2017). For each replicate, approximately 200 mg of freshly collected or flash-frozen 7-day-old leaves of AK58 were immediately chopped with a razor blade in ~1 ml of prechilled lysis buffer (15 mM Tris-HCl [pH 7.5], 20 mM NaCl, 80 mM KCl, 0.5 mM spermine, 5 mM 2-mercaptoethanol, 0.2% Triton X-100). The chopped slurry was filtered twice through Miracloth and once through a 40-μm filter. The crude nuclei were stained with DAPI and loaded into a flow cytometer (BD FACSCanto). Nuclei were purified by flow sorting and washed in accordance with previously reported methods (Lu et al., 2017). The sorted nuclei were incubated with 2 μl of Tn5 transposase (Vazyme, TD501) in 40 μl of tagmentation buffer (10 mM N-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid (TAPS)-NaOH [pH 8.0], 5 mM $MgCl_2$) at 37°C for 30 min without rotation. The integration products were purified using a Monarch DNA Cleanup Kit (T1030L, New England Biolabs) and then amplified using Phusion DNA polymerase for 10–13 cycles. The PCR cycles were as described previously (Buenrostro et al., 2013). Amplified libraries were purified with AMPure beads to remove primers.

Raw reads were trimmed with Trimmomatic v.0.36 (Bolger et al., 2014). Trimmed reads were aligned to the reference genome using Bowtie 2 v.2.2.4 (Langmead and Salzberg, 2012) with the following parameters: "bowtie2 -X 1000 –very-sensitive". Aligned reads were sorted using SAMtools v.1.3.1 (Li et al., 2009) and filtered by mapping quality (-q 30) to obtain unique reads, and clonal duplicates were removed using Picard v.2.16.0 (http://broadinstitute.github.io/picard/). The fraction of reads in peaks and a Venn diagram were used to reveal the consistency between replicates (Supplemental Table 12; Supplemental Figure 20). The genomic coverage was determined using deepTools (Ramirez et al., 2016) by counts per million (CPM) normalization with a bin size of 10 kb. Control data derived from naked DNA were used to avoid the aligning difference along the chromosome, and we used the ratio of normalized ATAC to the control in each bin to show chromatin accessibility. The same method was used to analyze ATAC-seq data of CS downloaded from the National Genomics Data Center (CRA003371; Yuan et al., 2022).

DATA AVAILABILITY

ChIP-seq data of G1812, Y2282, Zavitan, Svevo, CS, and AK58 and bisulfite sequencing data of WEW and DW were deposited in the NCBI Sequence Read Archive under BioProject number PRJNA622659 and the NGDC Genome Sequence Archive under BioProject number PRJCA014200. Other recently published datasets are available under the accession numbers listed in Supplemental Table 13.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

FUNDING

This research was supported by funding from the National Key Research and Development Program of China (2022YFF1003402), the China Natural Science Foundation (31371622), and the CAAS Innovation Program.

AUTHOR CONTRIBUTIONS

X.Z., J.J., and Z.R. designed the experiments. J.Z. and C.K. performed the ChIP-seq analysis. Y.X., J.Z., and Y.Z. performed the methylation analysis. Z.L. and J.Z. performed the ATAC-seq analysis. X.Z. and R.A. wrote the manuscript. D.C., Y.W., H.J., and Z.M. contributed to the revisions.

ACKNOWLEDGMENTS

We thank Profs. Weiwei Jin of China Agricultural University and Wenli Zhang of Nanjing Agricultural University for valuable discussions. No conflict of interest is declared.

Received: July 19, 2022
Revised: January 7, 2023
Accepted: February 1, 2023
Published: February 3, 2023

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. **215**:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science **357**:93–97. https://doi.org/10.1126/science.aan0032.

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.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods **10**:1213–1218. https://doi.org/10.1038/nmeth.2688.

Carbone, L., Nergadze, S.G., Magnani, E., Misceo, D., Francesca Cardone, M., Roberto, R., Bertoni, L., Attolini, C., Francesca Piras, M., de Jong, P., et al. (2006). Evolutionary movement of centromeres in horse, donkey, and zebra. Genomics **87**:777–782. https://doi.org/10.1016/j.ygeno.2005.11.012.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R., and Jiang, J. (2002). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14**:1691–1704. https://doi.org/10.1105/tpc.003079.

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

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinf. **9**:18. https://doi.org/10.1186/1471-2105-9-18.

Guo, X., Su, H., Shi, Q., Fu, S., Wang, J., Zhang, X., Hu, Z., and Han, F. (2016). De novo centromere formation and centromeric sequence expansion in wheat and its wide hybrids. PLoS Genet. **12**:e1005997. https://doi.org/10.1371/journal.pgen.1005997.

Han, Y., Zhang, Z., Liu, C., Liu, J., Huang, S., Jiang, J., and Jin, W. (2009). Centromere repositioning in cucurbit species: implication of the genomic impact from centromere activation and inactivation. Proc. Natl. Acad. Sci. USA **106**:14937–14941. https://doi.org/10.1073/pnas.0904833106.

Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. Science **293**:1098–1102. https://doi.org/10.1126/science.1062939.

Hunter, J.D. (2007). Matplotlib: a 2D graphics environment. Comput. Sci. Eng. **9**:90–95. https://doi.org/10.1109/Mcse.2007.55.

International Wheat Genome Sequencing Consortium (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science **361**. https://doi.org/10.1126/science.aar7191.

Jiang, J., Nasuda, S., Dong, F., Scherrer, C.W., Woo, S.S., Wing, R.A., Gill, B.S., and Ward, D.C. (1996). A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc. Natl. Acad. Sci. USA **93**:14210–14213. https://doi.org/10.1073/pnas.93.24.14210.

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

Karimi-Ashtiyani, R., Schubert, V., and Houben, A. (2021). Only the rye derived part of the 1BL/1RS hybrid centromere incorporates CENH3 of wheat. Front. Plant Sci. **12**:802222. https://doi.org/10.3389/fpls.2021.802222.

Kato, A., Albert, P.S., Vega, J.M., and Birchler, J.A. (2006). Sensitive fluorescence in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. Biotech. Histochem. **81**:71–78. https://doi.org/10.1080/10520290600643677.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods **12**:357–360. https://doi.org/10.1038/nmeth.3317.

Kishii, M., Nagaki, K., and Tsujimoto, H. (2001). A tandem repetitive sequence located in the centromeric region of common wheat (Triticum aestivum) chromosomes. Chromosome Res. **9**:417–428. https://doi.org/10.1023/a:1016739719421.

Koo, D.H., Sehgal, S.K., Friebe, B., and Gill, B.S. (2015). Structure and stability of telocentric chromosomes in wheat. PLoS One **10**:e0137747. https://doi.org/10.1371/journal.pone.0137747.

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.

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.

Lee, H.R., Zhang, W., Langdon, T., Jin, W., Yan, H., Cheng, Z., and Jiang, J. (2005). Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc. Natl. Acad. Sci. USA **102**:11793–11798. https://doi.org/10.1073/pnas.0503863102.

Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. **49**:W293–W296. https://doi.org/10.1093/nar/gkab301.

Li, B., Choulet, F., Heng, Y., Hao, W., Paux, E., Liu, Z., Yue, W., Jin, W., Feuillet, C., and Zhang, X. (2013). Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure. Plant J. **73**:952–965. https://doi.org/10.1111/tpj.12086.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, Z., Wang, M., Lin, K., 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.

Li, L.F., Zhang, Z.B., Wang, Z.H., Li, N., Sha, Y., Wang, X.F., Ding, N., Li, Y., Zhao, J., Wu, Y., et al. (2022). Genome sequences of five Sitopsis species of Aegilops and the origin of polyploid wheat B subgenome. Mol. Plant **15**:488–503. https://doi.org/10.1016/j.molp.2021.12.019.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**:923–930. https://doi.org/10.1093/bioinformatics/btt656.

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., Yue, W., Li, D., Wang, R.R.C., Kong, X., Lu, K., Wang, G., Dong, Y., Jin, W., and Zhang, X. (2008). Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres. Chromosoma **117**:445–456. https://doi.org/10.1007/s00412-008-0161-9.

Lo, A.W., Magliano, D.J., Sibson, M.C., Kalitsis, P., Craig, J.M., and Choo, K.H. (2001). A novel chromatin immunoprecipitation and array (CIA) analysis identifies a 460-kb CENP-A-binding neocentromere DNA. Genome Res. **11**:448–457. https://doi.org/10.1101/gr.gr-1676r.

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

Luo, M.C., Gu, Y.Q., Puiu, D., Wang, H., Twardziok, S.O., Deal, K.R., Huo, N., Zhu, T., Wang, L., Wang, Y., et al. (2017). Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature **551**:498–502. https://doi.org/10.1038/nature24486.

Lysak, M.A. (2014). Live and let die: centromere loss during evolution of plant chromosomes. New Phytol. **203**:1082–1089. https://doi.org/10.1111/nph.12885.

Ma, S., Wang, M., Wu, J., Guo, W., Chen, Y., Li, G., Wang, Y., Shi, W., Xia, G., Fu, D., et al. (2021). WheatOmics: a platform combining multiple omics data to accelerate functional genomics studies in wheat. Mol. Plant **14**:1965–1968. https://doi.org/10.1016/j.molp.2021.10.006.

Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M., Ormanbekova, D., Lux, T., Prade, V.M., Milner, S.G., et al. (2019). Durum wheat genome highlights past domestication signatures and future improvement targets. Nat. Genet. **51**:885–895. https://doi.org/10.1038/s41588-019-0381-3.

Mandáková, T., Hloušková, P., Koch, M.A., and Lysak, M.A. (2020). Genome evolution in Arabideae was marked by frequent centromere repositioning. Plant Cell **32**:650–665. https://doi.org/10.1105/tpc.19.00557.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. PLoS Comput. Biol. **14**:e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

Miller, J.T., Jackson, S.A., Nasuda, S., Gill, B.S., Wing, R.A., and Jiang, J. (1998). Cloning and characterization of a centromere-specific repetitive DNA element from Sorghum bicolor. Theor. Appl. Genet. **96**:832–839. https://doi.org/10.1007/s001220050809.

Montefalcone, G., Tempesta, S., Rocchi, M., and Archidiacono, N. (1999). Centromere repositioning. Genome Res. **9**:1184–1188. https://doi.org/10.1101/gr.9.12.1184.

Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S., and Jiang, J. (2003). Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. Genetics **163**:1221–1225.

Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R., and Jiang, J. (2004). Sequencing of a rice centromere uncovers active genes. Nat. Genet. **36**:138–145. https://doi.org/10.1038/ng1289.

Naish, M., Alonge, M., Wlodzimierz, P., Tock, A.J., Abramson, B.W., Schmücker, A., Mandáková, T., Jamge, B., Lambing, C., Kuo, P., et al. (2021). The genetic and epigenetic landscape of the Arabidopsis centromeres. Science **374**:eabi7489. https://doi.org/10.1126/science.abi7489.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. **20**:275. https://doi.org/10.1186/s13059-019-1905-y.

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.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. **44**:W160–W165. https://doi.org/10.1093/nar/gkw257.

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. Trends Genet. **16**:276–277. https://doi.org/10.1016/s0168-9525(00)02024-2.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139–140. https://doi.org/10.1093/bioinformatics/btp616.

Schmitz, R.J., Marand, A.P., Zhang, X., Mosher, R.A., Turck, F., Chen, X., Axtell, M.J., Zhong, X., Brady, S.M., Megraw, M., and Meyers, B.C. (2022). Quality control and evaluation of plant epigenomics data. Plant Cell **34**:503–513. https://doi.org/10.1093/plcell/koab255.

Schneider, K.L., Xie, Z., Wolfgruber, T.K., and Presting, G.G. (2016). Inbreeding drives maize centromere evolution. Proc. Natl. Acad. Sci. USA **113**:E987–E996. https://doi.org/10.1073/pnas.1522008113.

Su, H., Liu, Y., Liu, C., Shi, Q., Huang, Y., and Han, F. (2019). Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes. Plant Cell **31**:2035–2051. https://doi.org/10.1105/tpc.19.00133.

Sullivan, B.A., and Karpen, G.H. (2004). Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. Nat. Struct. Mol. Biol. **11**:1076–1083. https://doi.org/10.1038/nsmb845.

Tang, S., Tang, Z., Qiu, L., Yang, Z., Li, G., Lang, T., Zhu, W., Zhang, J., and Fu, S. (2018). Developing new oligo probes to distinguish specific chromosomal segments and the A, B, D genomes of wheat (Triticum aestivum L.) using ND-FISH. Front. Plant Sci. **9**:1104. https://doi.org/10.3389/fpls.2018.01104.

Ventura, M., Archidiacono, N., and Rocchi, M. (2001). Centromere emergence in evolution. Genome Res. **11**:595–599. https://doi.org/10.1101/gr.152101.

Ventura, M., Antonacci, F., Cardone, M.F., Stanyon, R., D'Addabbo, P., Cellamare, A., Sprague, L.J., Eichler, E.E., Archidiacono, N.,

and Rocchi, M. (2007). Evolutionary formation of new centromeres in macaque. Science **316**:243–246. https://doi.org/10.1126/science.1140615.

Ventura, M., Weigl, S., Carbone, L., Cardone, M.F., Misceo, D., Teti, M., D'Addabbo, P., Wandall, A., Björck, E., de Jong, P.J., et al. (2004). Recurrent sites for new centromere seeding. Genome Res. **14**:1696–1703. https://doi.org/10.1101/gr.2608804.

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez, R.H., Kolodziej, M.C., Delorean, E., Thambugala, D., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. Nature **588**:277–283. https://doi.org/10.1038/s41586-020-2961-x.

Wang, K., Wu, Y., Zhang, W., Dawe, R.K., and Jiang, J. (2014). Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Res. **24**:107–116. https://doi.org/10.1101/gr.160887.113.

Wicker, T., Matthews, D.E., and Keller, B. (2002). TREP: a database for Triticeae repetitive elements. Trends Plant Sci. **7**:561–562. https://doi.org/10.1016/S1360-1385(02)02372-5.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. **8**:973–982. https://doi.org/10.1038/nrg2165.

Wolfgruber, T.K., Sharma, A., Schneider, K.L., Albert, P.S., Koo, D.H., Shi, J., Gao, Z., Han, F., Lee, H., Xu, R., et al. (2009). Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. PLoS Genet. **5**:e1000743. https://doi.org/10.1371/journal.pgen.1000743.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. **35**:W265–W268. https://doi.org/10.1093/nar/gkm286.

Xue, C., Liu, G., Sun, S., Liu, X., Guo, R., Cheng, Z., Yu, H., Gu, M., Liu, K., Zhou, Y., et al. (2022). De novo centromere formation in pericentromeric region of rice chromosome 8. Plant J. **111**:859–871. https://doi.org/10.1111/tpj.15862.

Yan, H., Kikuchi, S., Neumann, P., Zhang, W., Wu, Y., Chen, F., and Jiang, J. (2010). Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. Plant J. **63**:353–365. https://doi.org/10.1111/j.1365-313X.2010.04246.x.

Yan, H., Jin, W., Nagaki, K., Tian, S., Ouyang, S., Buell, C.R., Talbert, P.B., Henikoff, S., and Jiang, J. (2005). Transcription and histone modifications in the recombination-free region spanning a rice centromere. Plant Cell **17**:3227–3238. https://doi.org/10.1105/tpc.105.037945.

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

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics **25**:1952–1958. https://doi.org/10.1093/bioinformatics/btp340.

Zhang, W., Lee, H.R., Koo, D.H., and Jiang, J. (2008). Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize. Plant Cell **20**:25–34. https://doi.org/10.1105/tpc.107.057083.

Zhao, H., Zeng, Z., Koo, D.H., Gill, B.S., Birchler, J.A., and Jiang, J. (2017). Recurrent establishment of de novo centromeres in the pericentromeric region of maize chromosome 3. Chromosome Res. **25**:299–311. https://doi.org/10.1007/s10577-017-9564-x.

Zhao, J., Hao, W., Tang, C., Yao, H., Li, B., Zheng, Q., Li, Z., and Zhang, X. (2019). Plasticity in Triticeae centromere DNA sequences: a wheat x tall wheatgrass (decaploid) model. Plant J. **100**:314–327. https://doi.org/10.1111/tpj.14444.

Zhong, C.X., Marshall, J.B., Topp, C., Mroczek, R., Kato, A., Nagaki, K., Birchler, J.A., Jiang, J., and Dawe, R.K. (2002). Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell **14**:2825–2836. https://doi.org/10.1105/tpc.006106.