PNAS RESEARCH ARTICLE | PLANT BIOLOGY

Non–B-form DNA tends to form in centromeric regions and has undergone changes in polyploid oat subgenomes

Qian Liu[a,b,1], Congyang Yi[a,b,1], Zeyan Zhang[a,b], Handong Su[c], Chang Liu[a,b], Yuhong Huang[a,b], Wei Li[d], Xiaojun Hu[e], Cheng Liu[f], James A. Birchler[g], Yang Liu[a,2], and Fangpu Han[a,b,2]

Edited by Yingxiang Wang, South China Agricultural University, China; received July 11, 2022; accepted November 23, 2022 by Editorial Board Member Pamela C. Ronald

Centromeres are the specialized regions of the chromosomes that direct faithful chromosome segregation during cell division. Despite their functional conservation, centromeres display features of rapidly evolving DNA and wide evolutionary diversity in size and organization. Previous work found that the noncanonical B-form DNA structures are abundant in the centromeres of several eukaryotic species with a possible implication for centromere specification. Thus far, systematic studies into the organization and function of non–B-form DNA in plants remain scarce. Here, we applied the oat system to investigate the role of non–B-form DNA in centromeres. We conducted chromatin immunoprecipitation sequencing using an antibody to the centromere-specific histone H3 variant (CENH3); this accurately positioned oat centromeres with different ploidy levels and identified a series of centromere-specific sequences including minisatellites and retrotransposons. To define genetic characteristics of oat centromeres, we surveyed the repeat sequences and found that dyad symmetries were abundant in oat centromeres and were predicted to form non–B-DNA structures in vivo. These structures including bent DNA, slipped DNA, Z-DNA, G-quadruplexes, and R-loops were prone to form within CENH3-binding regions. Dynamic conformational changes of predicted non–B-DNA occurred during the evolution from diploid to tetraploid to hexaploid oat. Furthermore, we applied the single-molecule technique of AFM and DNA:RNA immunoprecipitation with deep sequencing to validate R-loop enrichment in oat centromeres. Centromeric retrotransposons exhibited strong associations with R-loop formation. Taken together, our study elucidates the fundamental character of non–B-form DNA in the oat genome and reveals its potential role in centromeres.

polyploid oat | centromere | retrotransposon | non–B-form DNA | R-loop

Centromeres are the fundamental chromosomal regions responsible for accurate chromosome segregation during mitosis and meiosis (1–3). They provide the assembly sites for kinetochores that mediate interactions between the chromosomes and the spindle microtubules (4–6). Functional centromeres are marked epigenetically by the presence of the centromere-specific histone H3 variant (CENH3) (7–9), whose presence is both necessary and sufficient for centromere function. Despite their conserved function during chromosome segregation, centromeres are among the most dynamic and fast-evolving chromosomal regions (10). For most species, specific DNA sequences alone are neither sufficient nor necessary for centromere function supported by the existence of neocentromeres at a number of chromosomal sites (11). Thus, the functional significance of centromeric DNA in identity is unclear.

The DNA of eukaryotic cells is generally in a canonical right-handed B-form double helix (two DNA strands linked antiparallelly by forming base pairs in the B-form duplex). Interestingly, in addition to the canonical B-DNA structure, DNA can form noncanonical DNA structures such as G-quadruplexes, R-loops, hairpins/cruciforms, Z-DNA, triplexes, and slipped DNA (12–15). Noncanonical DNA structures, known as non–B-form DNA, have been implicated in gene expression (16, 17), telomere maintenance (18), double-strand breaks (19), and the life cycle of transposable elements (20). Recent studies highlighted a role for non–B-form DNA in centromere function (21–23). They found that centromeric DNA sequences from yeast and several animals are enriched in dyad symmetries, which show a predicted propensity to form non–B-form DNA. Furthermore, these secondary structures are directly visualized by single-molecule techniques, and in terms of the mechanism, CENP-B can modulate the topology of centromeric DNA, which is important for preserving centromere positioning and favoring centromere integrity in human cells (24). Whether plant centromeres fold into unique secondary DNA structures and the role of this non–B-form DNA in centromere remain unknown.

Significance

Centromeres of eukaryotic organisms are essential for faithful chromosome segregation. Non–B-DNA structures were recognized as important players in centromere activity and stability. To explore the organization and function of centromeric non–B-form DNA in polyploid oat, we accurately positioned oat centromeres with different ploidy levels and identified various centromeric repeats. Using computational programs, we found that predicted non–B-DNA structures were enriched at oat centromeres and potentially formed in centromeric repeats and CENH3 nucleosomes in vivo. Non–B-form DNA had undergone changes in polyploid oat subgenomes. Through applications of AFM and DRIP-seq, we validated R-loop enrichment in oat centromeres and observed R-loops derived from centromeric retrotransposons in vitro, suggesting the potential roles of non–B-DNA structures in plant centromeres.

The authors declare no competing interest.

This article is a PNAS Direct Submission. Y.W. is a guest editor invited by the Editorial Board.

[1]Q.L. and C.Y. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: yangliu@genetics.ac.cn or fphan@genetics.ac.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2211683120/-/DCSupplemental.

Published December 27, 2022.

Oat (*Avena sativa* L.) is one of the world's most important cereal crops owing to its use as an essential nutrient for both humans and animals. The genus *Avena* contains three major identified ploidy levels, that is, diploid, tetraploid, and hexaploid cytotypes (25, 26). The hexaploid oat (2n = 6× = 42, AACCDD) has been proposed to have originated from the formation of a CD genome tetraploid from C and D genome diploids, followed by hybridization to an A genome diploid, with whole-genome doubling at each stage to promote chromosome pairing (27, 28). Recently, the high-quality reference genomes of *Avena* species were released (27, 29); however, the landscape of centromeric repeats and their association with non–B-form DNA in oat have not been investigated in detail.

In this study, we identified functional centromeric regions of diploid, tetraploid, and hexaploid oat via chromatin immunoprecipitation sequencing (ChIP-seq) using anti-CENH3 antibody and isolated centromere-associated repeats as well as their positioning with CENH3 nucleosomes. We surveyed the predicted non–B-DNA structures and profiled a genome-wide identification of R-loops using DNA:RNA immunoprecipitation with deep sequencing (DRIP-seq). We discovered that non–B-form DNA is enriched in the oat centromeres and associated with centromeric transposable elements occupied with CENH3. These results provided additional insights into understanding non–B-form DNA in plant centromere specification and regulation.

Results

Identification of Functional Centromeres in Hexaploid Oat. To identify the location and sequence of functional centromeres of oat, we developed a CENH3 antibody that could recognize the oat centromeres of all chromosomes (*SI Appendix*, Fig. S1 *A* and *B*). ChIP-seq with CENH3 antibody was performed using nuclei isolated from leaf tissue of hexaploid oat (ACD, 2n = 2× = 42). To confirm the specificity of ChIP experiments, we use fluorescence in situ hybridization (FISH) of ChIPed DNA on metaphase chromosomes of hexaploid oat. The results showed strong signals at centromeres of each chromosome, supporting the enrichment of centromeric sequences (*SI Appendix*, Fig. S1*C*). The ChIPed DNA and input control DNA were then sequenced using the NovaSeq platform, both generating approximately 18 million 150-bp paired-end reads. The filtered reads were mapped to the high-quality reference genome of Sang oat, which has been recently sequenced using the Oxford Nanopore ultralong sequencing and Hi-C technologies (29). Approximately, 78% of the reads were aligned to a unique position in the reference genome (*SI Appendix*, Table S1).

The unique ChIP-seq reads that were normalized by input were plotted in 100-kb windows along the chromosome position (Fig. 1). In contrast to the discontinuous CENH3 peaks that were detected in some chromosomes of wheat and switchgrass (30, 31), significant signal sequence enrichment was observed in the centromeric regions of the 21 oat chromosomes, indicating a high-quality assembly of oat centromeres (Fig. 1). The size of the core region of CENH3 binding in hexaploid oat centromeres varies from 5.0 to 9.8 Mb, with an average centromere size of ~6.9 Mb (Table 1), which is similar to that of wheat centromeres. Of the three subgenomes, the AA subgenome showed the largest centromere size, followed by the DD and CC subgenomes.

Predicted Non–B-Form DNA Tends to Form in Oat Centromeric Regions. The high-quality oat genome and precise knowledge of its centromere location provide an opportunity to inquire into the genetic characteristics of the oat centromere. Previous work found that non–B-form DNA is enriched in human centromeres and has been implicated in regulating its activity and stability (23). To investigate whether oat centromeres fold into unique secondary DNA structures, we first examined the distribution of dyad symmetries in oat centromeres with the consideration that dyad symmetries are prone to adopt non–B-DNA conformations. We applied the computational EMBOSS palindrome algorithm to identify dyad symmetries <10-bp DNA based on the genome sequence. The length of most dyad symmetries varies in the range of 4 to 6 bp. We compared the density of dyad symmetries in the centromere, the pericentromere, and the chromosome arms of the three subgenomes. We found that in all the three subgenomes, centromeres displayed higher dyad symmetry densities than chromosome arms (Mann–Whitney *U* test, $P < 0.05$) (Fig. 2*A*), suggesting that dyad symmetries are abundant in oat centromeres. RNAfold was previously used to predict the folding free energies of non–B-DNA structures, a more negative folding free energy indicates more stable non–B-DNA structures (21). Indeed, we found that centromeric DNA tends to form more stable secondary structures (Fig. 2*B*). This phenomenon was even more pronounced in the centromeres of the AA and DD subgenomes but less obvious in the centromeres of the CC subgenome. Furthermore, the folding free energies were negatively associated with the density and length of palindromes in the centromeres (*SI Appendix*, Fig. S2 *A* and *B*).

To determine whether centromeres are susceptible toward adopting non–B-DNA structures, we used the non–B-DNA motif search tool (nBMST) (32) that can generate accurate predictions of six different non–B-DNA-forming motifs: A-phased DNA repeats (bent DNA), direct repeats (slipped structures), mirror repeats (triplex DNA), inverted repeats (cruciform structures), alternating purine–pyrimidine tracts (Z-DNA), and G4 motifs (G-quadruplexes) (32). When we plotted the number of predicted non–B-form DNA in a 500-kb window along the entire oat genome, we observed that Z-DNA, slipped DNA, and DNA bending tended to form in the centromeres (Fig. 2*C* and *SI Appendix*, Figs. S3 and S4). Permutation tests also revealed these three non–B-form DNA are highly enriched in the oat centromeres, while the other four non–B-form DNA did not show this tendency (Fig. 2*D*). Taken together, these results suggest dyad symmetries are enriched in oat centromeric regions and may fold into different non–B-DNA conformations.

Centromeric Repeats Contribute to the Formation of Non–B-Form DNA. The centromeric region is mainly composed of transposable elements, and the results that non–B-form DNA is enriched in centromeres prompted us to investigate which types of centromeric repeats of *A. sativa* (CRA) contribute to the formation of non–B-DNA structures. We first constructed the de novo repeat clusters analysis for hexaploid oat with RepeatExplorer software using 1.2 million randomly selected sequence reads from the input. Then, the CENH3 ChIP-seq reads and input reads were mapped to these repeat clusters to calculate the ratios of each cluster. The ChIP to input ratio was indicative of the relative enrichment of each repeat in the centromeres. Of these clusters, 64 clusters showed a ratio greater than threefold (Fig. 3*A*), suggesting that these repeat clusters were likely associated with the centromeres. Using the Basic Local Alignment Search Tool (BLAST), we found six clusters were specifically mapped to the centromeres of the AA and DD subgenomes, seven clusters were mapped to the centromeres of the CC subgenome, and 51 clusters span the entire functional core of all the oat centromeres. All the centromeric repeats belonged to long terminal repeat (LTR) retrotransposons of the *Gypsy*- and *Copia*-like families (Dataset S1). Four clusters were selected for

further cytological confirmation. FISH assays showed that clusters CRA39, CRA120, and CRA299 produced clear signals in all the oat centromeres, and the cluster CRA82 was only abundant in centromeric regions of the AA and DD subgenomes (Fig. 3*C*). We did not identify tandem repeats in this analysis. Thus, we used another software, tandem repeats finder (33), to investigate whether some tandem repeats were enriched in oat centromeres. Four minisatellites showed fivefold enrichment and FISH detected one pair of strong centromere signals and two pairs of weak centromere signals for minisatellite 1 (M1) and M2 while M3 and M4 located to five pairs of centromeric regions (Fig. 3 *B* and *C*).

To determine which types of centromeric repeats are contributing most to the likelihood of adopting non–B-DNA structures, we first computed the content of the repetitive elements in the genome as a control. Then, fold enrichment was calculated by comparing the percentage of six types of predicted non–B-DNA conformations in repetitive elements and control sequences. Extensive non–B-form DNA variation was observed between transposon families. For example, Z-DNA was more likely to form in *SINE* and *Gypsy* retroelements; triplex DNA tended to form in transposons, such as *CACTA*, *Mutator*, *hAT*, *Helitron*, and *Tc1-Mariner*; cruciform DNA was overrepresented in *hAT*, *Tc1-Mariner*, *Copia* (belonging to the LTRs), and *PIF-Harbinger*; G-quadruplex was highly aggregated in *hAT*, *Tc1-Mariner*, *CACTA*, and *PIF-Harbinger*; slipped DNA showed significant positive overlap with *CACTA*, *Mutator*, and *Helitron*; the majority of bent DNA preferentially form in *Gypsy* (Fig. 3*D*). Notably, recent studies have shown that centromeric satellites are strongly associated with non–B-DNA structures (21–23). Therefore, these results suggest that centromeric repetitive elements are enriched for predicted non–B-DNA structures; more than one repeat class or family at centromeres contributes to the enrichment of individual non–B-DNA structures, indicating that centromeric repeats may differently contribute to non–B-DNA structure formation.

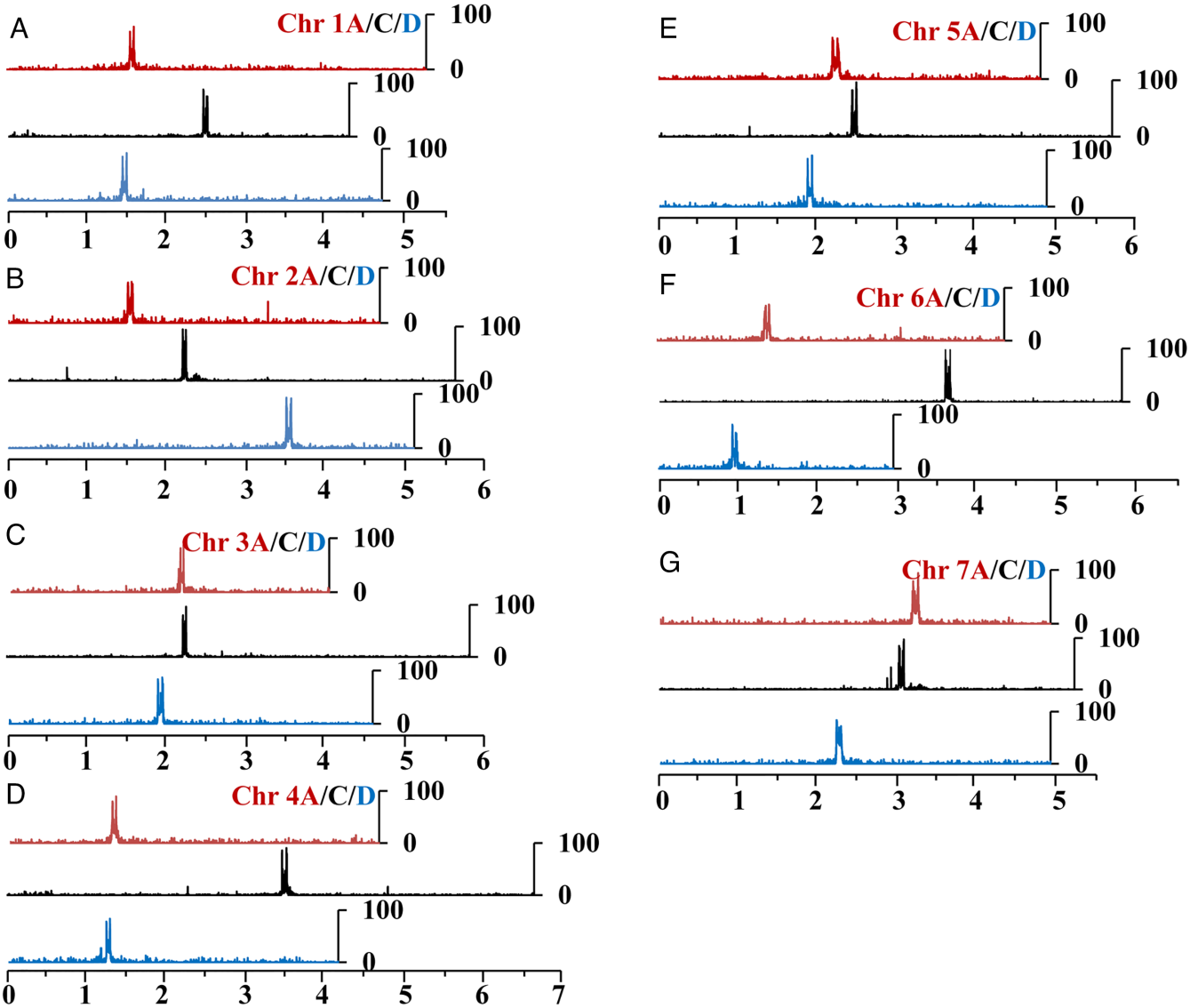

Fig. 1. Genome-wide mapping of CENH3 ChIP-seq reads to the oat reference genome (AACCDD). (*A*–*G*) Schematic overview of CENH3 ChIP-seq reads of hexaploid oat normalized by input on homoeologous group chromosomes 1 to 7 (RPGC normalization in 5-kb bins). The mapped read density was plotted in 100-kb windows along 21 chromosomes. The X axis shows the positions (×100-Mb) on each chromosome. The Y axis represents the normalized sequence read count ratio between ChIP-seq and input on each position, which was calculated based on the ratio of reads per million/100. The different colored lines represent the read density in different oat subgenomes (red, AA subgenome; black, CC subgenome; and blue, DD subgenome).

Table 1. Positions of functional CENH3-binding regions in hexaploid oat centromeres

Chr	Chr size (Mb)	Cen location (Mb)	Cen size (Mb)	Chr	Chr size (Mb)	Cen location (Mb)	Cen size (Mb)	Chr	Chr size (Mb)	Cen location (Mb)	Cen size (Mb)
1A	525	153.8–161.3	7.5	1C	426	243.5–249.5	6.0	1D	469	143.2–150.0	6.8
2A	469	150.0–158.0	8.0	2C	538	209.5–214.5	5.0	2D	516	353.0–361.0	8.0
3A	409	226.5–231.5	5.0	3C	574	218.5–223.5	5.0	3D	454	185.0–192.5	7.5
4A	460	126.0–133.0	7.0	4C	667	348.0–355.0	7.0	4D	415	122.5–128.5	6.0
5A	476	216.0–225.8	9.8	5C	573	243.5–250.5	7.0	5D	489	186.6–194.0	7.4
6A	445	133.0–140.0	7.0	6C	579	357.5–364.5	7.0	6D	298	90.5–97.0	6.5
7A	490	318.0–327.5	9.5	7C	506	292.0–298.6	5.4	7D	490	220.6–228.5	7.9

Cen, centromere; Chr, chromosome.

R-Loops Are Common Features of Oat Centromeres. As one type of non–B-DNA conformations, R-loops are accumulated in centromeric regions in human, *Arabidopsis*, and maize to promote faithful chromosome segregation and chromatin loop formation (34, 35). To answer whether R-loops are prevalent in oat centromeres, we used an DRIP-seq approach to enable the genome-wide detection of R-loops in hexaploid oat (negative control treatment with RNase H) (*SI Appendix*, Fig. S5). We detected 954,788 Watson DNA strand–related R-loop (wR-loop) peaks and 952,020 Crick DNA strand–related R-loop (cR-loop) peaks (*SI Appendix*, Fig. S6*A*), which covered ~5% of the oat genome (*SI Appendix*, Fig. S6*B*). Most of the peaks were 200 to 1,000 base pairs (bp) long (*SI Appendix*, Fig. S6*C*). A GA-rich motif of R-loops was identified by using MEME-ChIP (36) (*SI Appendix*, Fig. S6*D*). Oat sense R-loops mainly formed in both promoters and transcription termination sites, while antisense R-loops occurred around the transcription start sites compared with flanking regions (*SI Appendix*, Fig. S7). Furthermore, they tend to form in regions of high GC and AT skews (*SI Appendix*, Fig. S8). Altogether, the sequence signatures and distribution pattern of R-loops shared a conserved signature across species. Next, we used MACS2 (37) to call R-loop peaks and profiled the distribution of R-loop peaks along each chromosome. We found a chromosome-level trend of R-loops increasing toward pericentromeres and centromeres (Fig. 4*A*). We also performed immunostaining of anti-DNA–RNA hybrid antibody (S9.6 antibody) combined with FISH using the probe of centromere-specific retrotransposon (CRA39). Many R-loop signals were scattered along the chromosomes, and a few signals were observed in proximity of CRA39 foci (Fig. 4*B*). Our previous studies in maize have shown that R-loops were strongly accumulated at the 45S rDNA region. Similarly, very prominent R-loop signals were also observed in the 45S rDNA region of oat (Fig. 4*B*). However, it seems that there is no correlation with any feature on the heat maps (Fig. 4 *A* and *B*). To assess the assembly quality of the rDNA region, the full-length 45S rDNA was used as a query for BLAST analysis to the oat reference genome. As expected, most of the sequence was aligned to one unassigned scaffold, indicating that 45S rDNA might not be fully assembled in the oat reference genome (*SI Appendix*, Fig. S8*D*). Thus, the 45S rDNA region is presented quite small in the graph (Fig. 4*A*). Taken together, our bioinformatic and experimental results suggested that R-loops were enriched in oat centromeric and pericentromeric regions.

To determine which structural elements in CRA are the major contributions for R-loop formation, we profiled the distribution of R-loops on CRA39, CRA82, CRA120, and CRA299 (Fig. 4*C* and *SI Appendix*, Fig. S9 *A* and *B*). We found that R-loops localized to both ends of the retrotransposons (<2-kb). To get a better insight into the architecture of R-loops formed at these retrotransposons, we used an in vitro transcription system to characterize the consequences of R-loop formation on the CRA39 and CRA82 templates. As reported previously (35), R-loop formation on circular templates was associated with a pronounced shift in plasmid mobility on agarose gels. RNase H treatment completely reverted this upward shift (Fig. 4*D*). Dot blots with the anti-S9.6 antibody also validated the formation of stable R-loops at CRA39 and CRA82 (Fig. 4*E*). Furthermore, we used atomic force microscopy (AFM) to directly visualize R-loop structures on the CRA39 and CRA82 templates. According to the method described for R-loop imaging of maize retrotransposons (38), we unambiguously identified R-loop structures on the CRA39 and CRA82 templates due to its displaced single-stranded DNA. We observed that R-loops were positioned at 35% (±5%) from one end of DNA, consistent with the expected location of the R-loops on the full-length CRA39 and CRA82 (Fig. 4*F* and *SI Appendix*, Fig. S10), which is consistent with the distribution of DRIP-seq reads on CRA39 and CRA82. In contrast, we did not observe an R-loop structure in the CRA control in AFM. Taken together, these results from agarose gels, S9.6 dot blots, and AFM validation showed that R-loops formed on the oat centromeric retrotransposons.

Non–B-DNA Structures Are Predicted to Form in CENH3-Binding Regions. Active centromeres are marked by nucleosomes assembled with CENH3. However, it remains unknown how the newly synthesized CENH3 is incorporated into centromeres. Previous results revealed that highly repeating centromere sequences were linked to centromere locations or CENH3 binding in rice, maize, wheat, and other species (31, 39–41), indicating their potential roles in centromere function. Thus, to investigate whether the identified centromeric retrotransposons from oat contribute to the assembly of CENH3 nucleosomes, we surveyed the association between CENH3 nucleosomes and retrotransposons, including CRA39, CRA82, CRA120, and CRA299. To do this, we followed a similar approach used in analyzing human and wheat centromeres (31, 42, 43). The CENH3-ChIPed and Input-seq reads represent the CENH3 and canonical nucleosome fragments, respectively. The joined reads were aligned to the retrotransposon sequences using BWA-MEM software (44). The midpoint distribution of the fragments was treated as the location of the sequence on the CENH3 or canonical nucleosomes. Five major binding peaks were identified on CRA39 for CENH3-ChIP-seq reads, indicating phasing of CENH3 nucleosomes on CRA39. Moreover, the input nucleosomes were present at the same positions on CRA39, suggesting that most of the CRA39 sequences for each repeat were wrapped around CENH3 nucleosomes (Fig. 5*A*). Similarly,

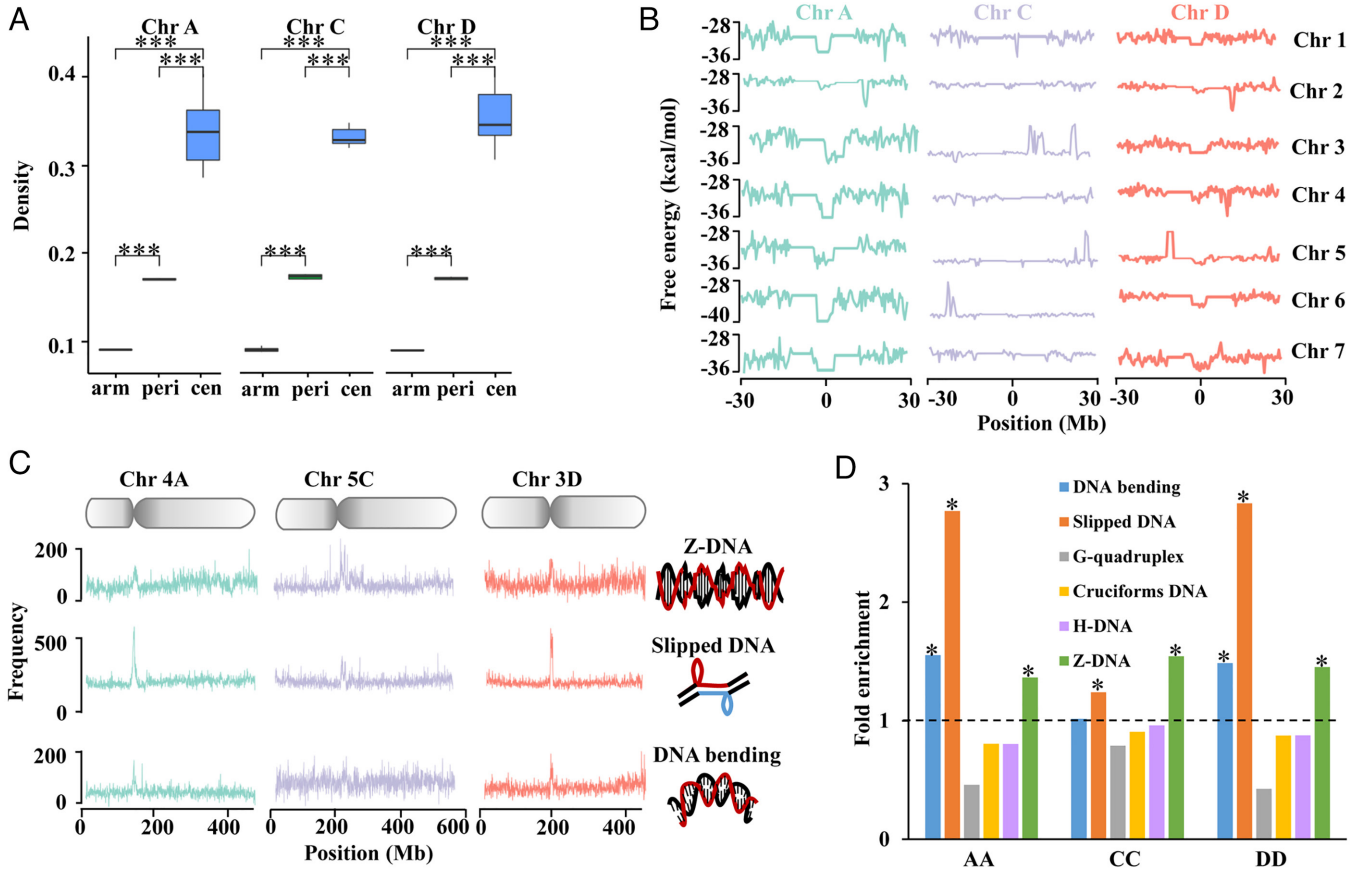

Fig. 2. Predicted non–B-form DNA is enriched in oat centromeres of three subgenomes. (*A*) Dyad symmetry density at oat centromeres relative to pericentromeres and chromosome arms (total number of palindromic positions normalized by sequence length). Dyad symmetries of centromeres showed a significantly higher density (Mann–Whitney *U* test, $P < 0.05$). Asterisks denote the observed value was greater than 90% of the Mann–Whitney *U* test. Arm, chromosome arm; peri, pericentromere; and cen, centromere. (*B*) Predicted DNA secondary structure folding free energy along centromeres and 30-Mb up- and downstream regions. The cyan, purple, and orange lines represent DNA secondary structure folding free energy predictions in 500-kb windows along the chromosomes of AA, CC, and DD subgenomes, respectively. The X axis represents the position of DNA dyad symmetries on the chromosome; the Y axis represents the folding free energy. (*C*) Diagram of frequency of different non–B-DNA structures on representative chromosomes (Chr 4A, Chr 5C, and Chr 3D). The X axis represents the position of DNA dyad symmetries along the chromosomes; the Y axis represents the frequency of non–B-form DNA in 500-kb window along the chromosomes. The different colored lines represent different oat subgenomes (cyan, AA subgenome; purple, CC subgenome; and orange, DD subgenome). Schematic diagrams of different types of non–B-DNA structures including Z-DNA, slipped DNA, and DNA bending (*Right*). (*D*) The fold enrichment of different types of centromeric non–B-form DNA relative to the whole AA, CC, and DD subgenomes. The dashed line represents enrichment fold = 1.0 (permutation test; asterisks denote the observed value was >90% of the permutation value). The different colored lines represent different types of non–B-form DNA (blue, DNA bending; orange, slipped DNA; gray, G-quadruplex; yellow, cruciform DNA; purple, H-DNA; and green, Z-DNA).

several major peaks were also observed on CRA82, CRA120, and CRA299, which implies the presence of a specific position for mapping CENH3 nucleosomes (Fig. 5*B* and *SI Appendix*, Fig. S9 *D* and *E*). These results indicated that CENH3 nucleosomes are phased with specific sites on the oat centromeric retrotransposons.

Considering that centromeric repeats are capable of selectively forming a wide variety of non–B-DNA structures (Fig. 3*D*), we asked whether non–B-form DNA is associated with CENH3 nucleosomes. To this end, we used two different methods to analyze the potential correlation of the predicted non–B-form DNA with CENH3 nucleosomes. First, we determined the positions of CENH3 nucleosomes from CENH3-ChIPed reads using nucleR (45) and evaluated overlap of six types of non–B-DNA conformations and CENH3 nucleosomes. Although cruciform DNA is not enriched at centromeres (*SI Appendix*, Fig. S4 *B* and *D*), it overlaps with CENH3 nucleosomes significantly more than expected. Additionally, CENH3 nucleosomes showed the high degree of overlap with DNA bending, slipped DNA, Z-DNA, and triplex DNA. The lowest overlap was observed between CENH3 nucleosomes and G-quadruplexes (Fig. 5*C*). Second, we profiled the distribution of CENH3-ChIP reads around predicted non–B-form DNA in centromeric regions; the results indicated that the formation regions of predicted slipped DNA, Z-DNA, DNA bending, cruciform DNA, and G-quadruplexes were enriched in CENH3 (Fig. 5*D*). However, no CENH3 enrichment was detected for predicted triplex DNA (*SI Appendix*, Fig. S9). Our previous work showed that R-loops tend to form in CENH3 nucleosomes in *Arabidopsis* and maize (35, 38). To assess whether the colocalization pattern was conserved in oat, we analyzed the association between R-loop peaks and CENH3 nucleosomes using CENH3-ChIP and R-loop data and found that R-loops were also highly colocalized with CENH3 nucleosomes (Fig. 5 *E* and *F*). Altogether, these results provided clues that non–B-form DNA may play a role in CENH3 nucleosome identity and assembly by acting as a "docking zone."

Non–B-DNA Conformations Are Involved in the Rearrangement of Centromeres during Evolution. Polyploidization frequently happens in plant evolution and is usually associated with rapid structural and functional alterations of genomes (31); centromeres are the most rapidly evolving genomic regions and vary dramatically in size and organization. Non–B-DNA structures can

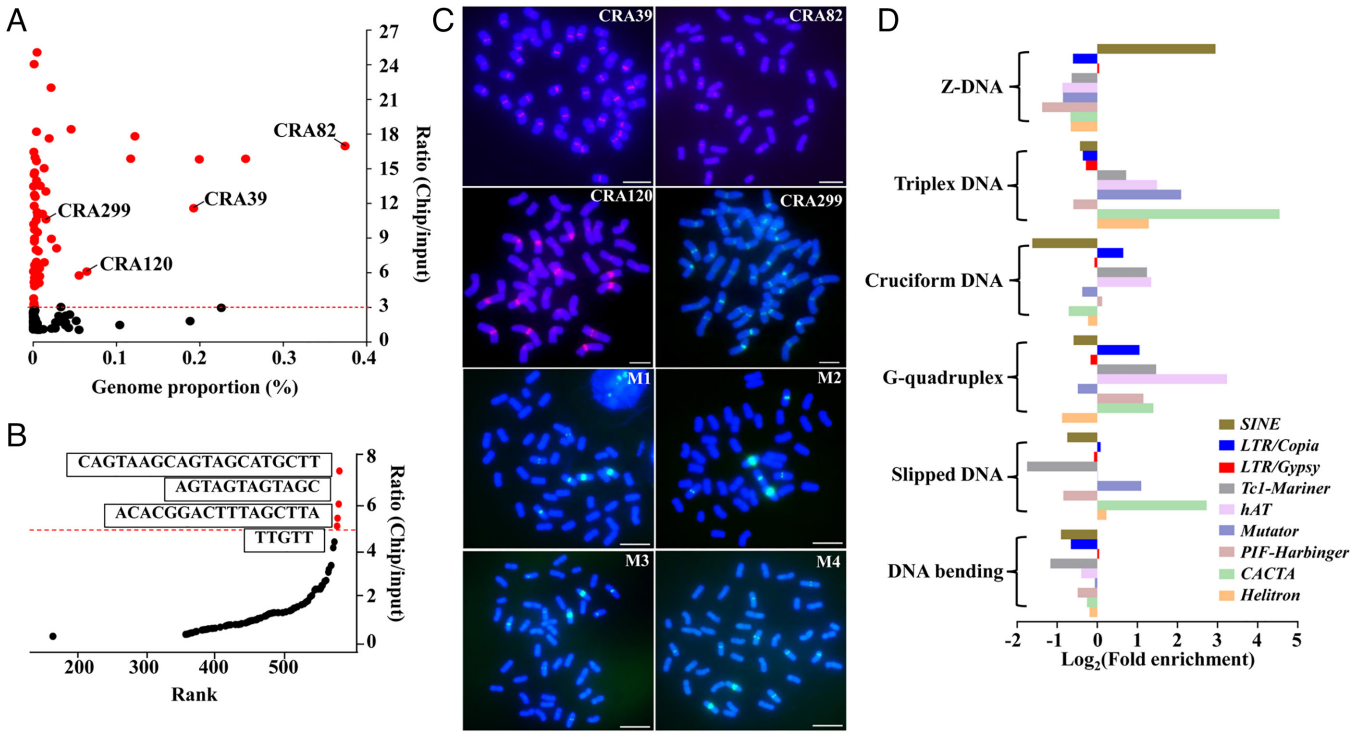

Fig. 3. Centromeric repeats contribute to the formation of non–B-form DNA. (*A*) Identification of the repeat clusters in hexaploid oat centromeres. Repeat clusters are represented by dots. The X axis and Y axis represent the genome proportion and the relative enrichment of Chip-seq reads of each cluster, respectively. Hundred abundant repeat clusters were shown (black and red dots). (*B*) Identification of minisatellites in centromeres of hexaploid oat. Minisatellites are represented by dots and ranked by their confidence. The Y axis represents the relative enrichment of Chip-seq reads of each minisatellite. Red dots indicate the minisatellites selected for investigation in this study. (*C*) FISH confirms the centromere location of four retrotransposons and minisatellites in hexaploid oat. Repeat clusters CRA39 (red), CRA82 (red), CRA120 (red), and CRA299 (green); minisatellite M1(green), M2 (green), M3 (green), and M4 (green). Blue indicates chromosomes counterstained with DAPI. (Scale bar, 10 μm.) (*D*) Bar graphs showing $\log_2$-fold enrichment of different non–B-form DNAs in indicated TEs. The X axis corresponds to the $\log_2$-transformed fold enrichment (permutation test; asterisks denote the observed value was >90% of the permutation value). *SINE* (blue), *LTR/Copia* (blue), *LTR/Gypsy* (red), *Tc1-Mariner* (gray), *hAT* (pink), *Mutator* (purple), *PIF-Harbinger* (fuchsia), *CACTA* (green), and *Helitron* (yellow).

affect DNA synthesis and lead to genome instability. To further investigate the changes of non–B-form DNA that occurred within the centromeric regions during oat evolution, we first identified centromeric regions of diploid (*Avena atlantica,* AA, 2n = 14; *Avena eriantha,* CC, 2n = 14) and tetraploid (*Avena maroccana*, CCDD, 2n = 4× = 28) relatives by ChIP-seq with anti-CENH3 antibody. The trimmed reads were mapped to a unique position in the reference genomes of *A. atlantica* (27), *A. eriantha* (27), *and A. insularis* (CCDD, 2n = 4× = 28) (46), respectively. The distribution of ChIP-seq reads normalized by input was displayed in 500-kb windows along the chromosomes. Significant sequence enrichment was observed in the centromeres of diploid and tetraploid oat (Fig. 6*A* and *SI Appendix*, Figs. S11–S15). Then, we used defined criteria to delimit the boundaries of each centromere and determined the sizes of various centromeres. The average size of CENH3-enriched centromere cores of *A. atlantica*, *A. eriantha*, and *A. maroccana* was 5.6, 6.3, and 5.4 Mb, respectively (*SI Appendix*, Tables S2 and S3).

Sequence variation is tightly associated with non–B-DNA conformations and linked to centromere location or CENH3 binding. Previous studies found that centromere positions might vary in wheat genomes at different ploidy levels (31). To assay for such variation in oat, we remapped the CENH3-ChIP-seq reads from *A. atlantica*, *A. eriantha*, and *A. insularis* to the reference genome of hexaploid oat. This approach makes it possible to identify conserved and varied centromeric regions. Approximately 36.9%, 38%, and 40% of the reads were aligned to distinct positions in the AA, CC, and CCDD chromosomes of the reference genome, respectively (Fig. 6*B*). Significant peaks indicating the location of the centromeres were observed on all the chromosomes. However, the size and position of centromeres vary widely, especially from the progenitor of CC and tetraploid (CCDD) to hexaploid oat (Fig. 6*B*). These results suggest that rapid divergence of centromeric DNA occurs from diploid to tetraploid to hexaploid oat during and/or after polyploidization.

As the centromeres of *A. atlantica*, *A. eriantha*, and *A. insularis* are well assembled, this provides an unprecedented opportunity to investigate the structural variation by collinearity analysis in oat centromeres during polyploidization. The synteny map revealed extensive structural rearrangements in the centromeric and pericentromeric regions between *A. atlantica* and Sang in the AA subgenome, *A. eriantha*, *A. insularis*, and Sang in the CC subgenome, and *A. insularis* and Sang in the DD subgenome (Fig. 6*C* and *SI Appendix*, Fig. S16). For the subgenomes AA and DD, small-scale inversions were detected between orthologous chromosomes. Genes in Cen4A and Cen5A of *A. atlantica* showed no collinearity with the AA subgenomes of Sang, indicating gene loss in Sang Cen4A and Cen5A. More strikingly, we observed an excess of syntenic gene loss at and near the centromeric regions of *A. eriantha* (*SI Appendix*, Fig. S16). These results suggest that large chromosomal deletions occurred from the progenitor of the AA and CC subgenomes to hexaploid oat. Taken together, the detection of gene loss and inversions indicates that oat centromeres have undergone rearrangement during evolution, which also led us to suggest that we may observe non-B conformational variation in oat centromeres at different ploidy levels.

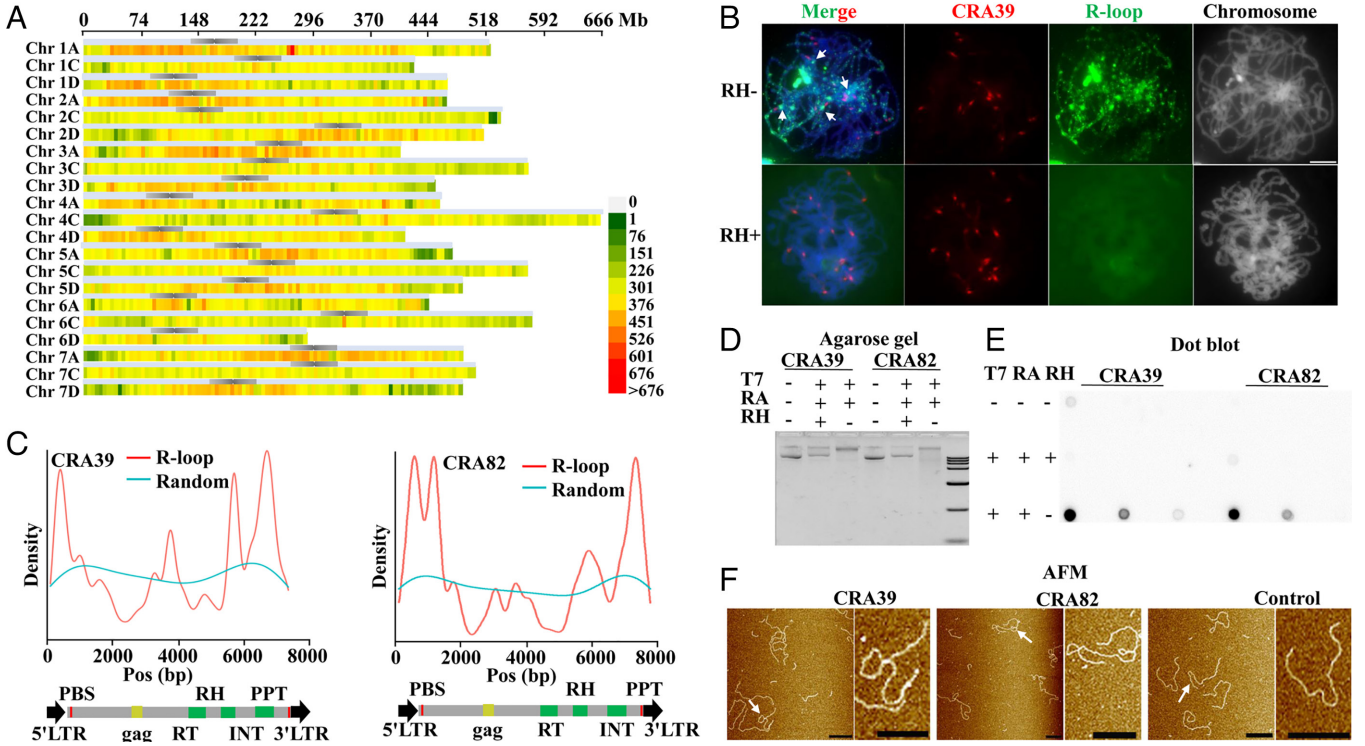

Fig. 4. R-loops are enriched in oat centromeres and pericentromeres. (*A*) Heat map of R-loop peaks on the whole chromosomes of oat. The X axis represents positions on chromosomes. Schematic diagram of chromosomes with pericentromere shade of gray color (*Top*). (*B*) R-loop (green) and CRA39 (red) signals on chromosomes during the pachytene of meiosis I. RNase H treatment (RH+) abolished the R-loop signals on the chromosomes. (Scale bar, 10 μm.) (*C*) Distribution of R-loops peaks on CRA39 and CRA82. The red line represents the distribution of R-loops. The 10,000 randomly selected 150-bp reads were used as a control (light blue). Schematic drawing of sequence (with domains highlighted) for CRA39 and CRA82. RT, reverse transcriptase; RH, RNase H; and INT, integrase. (*D* and *E*) The circular plasmids pET-30a-CRA39/CRA82 were transcribed in vitro; negative control was treated with both RNase A and RNase H as indicated. The purified DNA was run on agarose gels (*D*) or used to perform dot blot analysis with anti-S9.6 antibody onto the membrane (*E*). (*F*) R-loops transcribed from pET-30a-CRA39/CRA82/control were visualized using AFM. Arrows show the magnifications of the molecules identified. (Scale bar, 200 nm.)

To further evaluate the changes of predicted non–B-form DNA that occurred within centromeric regions during oat evolution, we calculated the enrichment of six types of non–B-DNA conformations in centromeric regions of diploid (AA and CC) and tetraploid (CCDD) oat. The results showed that predicted Z-DNA, slipped DNA, and DNA bending were enriched in centromeres of the AA diploid and D subgenome of tetraploid oat, which were consistent with the results of hexaploid oat (Fig. 6*D*). However, bent DNA was not enriched in the centromeres of CC diploid and C subgenome of tetraploid oat. Next, we compared the proportions of the six types of predicted non–B-form DNA in the centromeres of polyploid oat subgenomes. Predicted non–B-form DNA showed distinct patterns in the A, C, and D subgenomes from diploid to tetraploid to hexaploid oat. For example, a greater fraction of H-DNA occurs in the centromeres of AA diploid (~23%), while it was depleted in the A subgenome of hexaploid oat (~7%); cruciform DNA and Z-DNA increased in oat centromeres during the evolution from diploid to tetraploid to hexaploid oat (Fig. 6*E*). These results suggested that frequent chromosomal rearrangements occurred during the evolution of oat of non–B-form DNA exists.

Discussion

In most higher eukaryotes, centromeres are composed of satellite repeats and retrotransposons. Due to the very large and complex genome, cytogenetic and genomic analysis of oat centromeres lag far behind other major crops. Although genome-wide repetitive DNA in oat had been identified based on the unprocessed oat genomic DNA (47), the genetic and epigenetic characteristics of oat centromeres have not been reported, partly due to the challenges of centromere sequences assembly (27). Here, we report CENH3 ChIP-seq analysis in polyploid oat and used a graph-based clustering approach to identify repeat DNA sequences associated with centromeres. Different from the DNA composition of centromeres reported in maize (48), *Arabidopsis* (49), rice (50), and switchgrass (30), we identified abundant centromeric repeats originating from *LTR/Gypsy* retrotransposons (Dataset S1). However, we did not detect centromere-specific tandem repeats. The presence of chromosome-specific satellite-free centromeres was also reported in brassica, cotton, potato, wheat, and chicken (31, 51–55). It is possible that the centromeres in both the polyploid and diploid oat are still undergoing dynamic changes. CENH3-containing nucleosomes exhibit a strong preference for the young repeats (56); young retrotransposons in oat centromeres are expected to evolve, potentially, leading to the accumulation of satellite arrays. Another possibility is that satellite repeats were originally present in oat core centromeres, but they were excluded from the active centromeric regions by subsequent retrotransposon insertions into the CENH3 regions (57). The future analysis of centromeric sequence composition in oat progenitors will provide insight into sequence variation during centromere evolution. Furthermore, in order to determine whether these retrotransposons are functional in centromeres, we analyzed the binding sites of CENH3 or bulk nucleosomes on the retrotransposons and found that these retrotransposons have the capacity to bind CENH3 (Fig. 5 and *SI Appendix*, Fig. S9 *D* and *E*). However, these retrotransposons share no similarity.

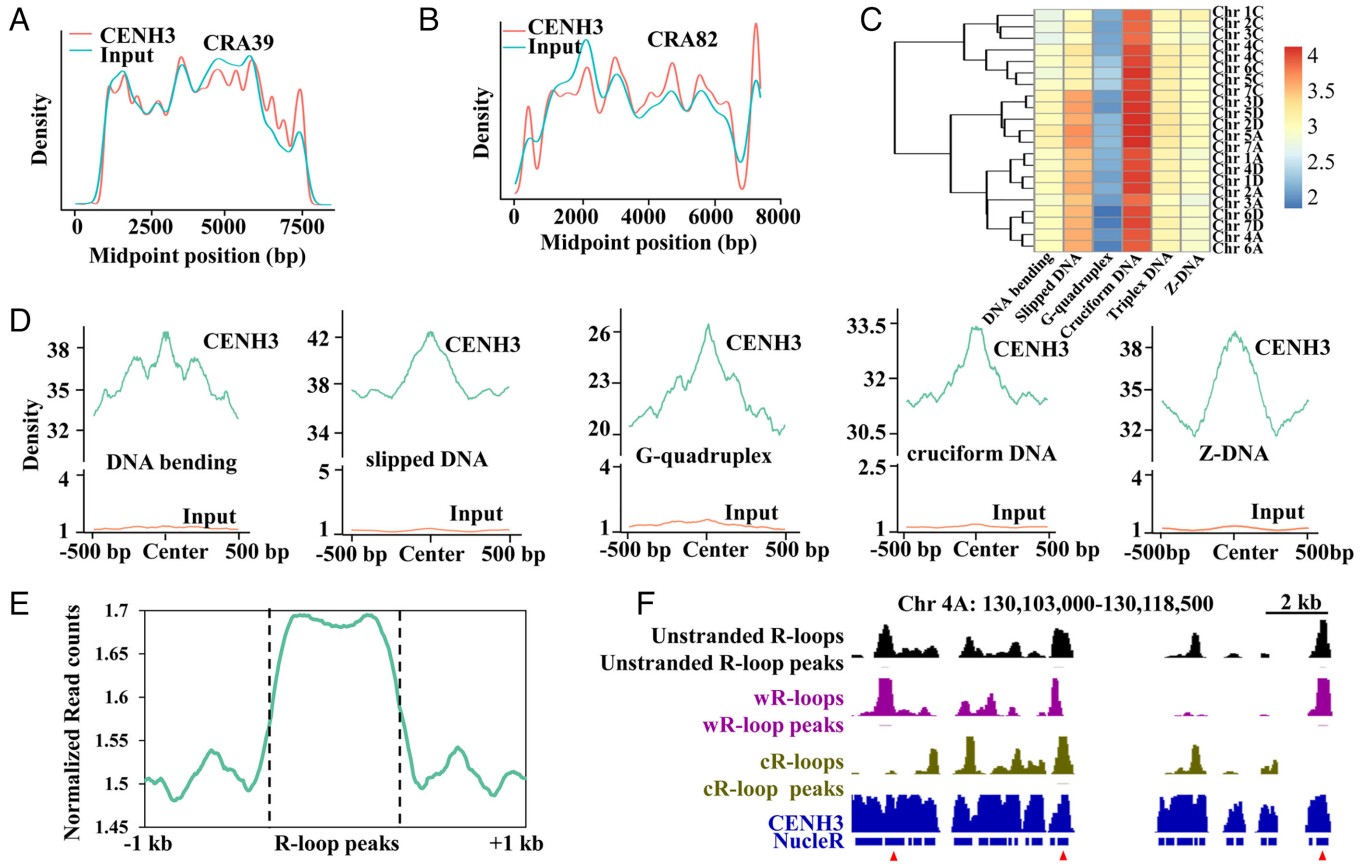

Fig. 5. Non–B-DNA structures tend to form in CENH3-binding regions. (*A* and *B*) Distribution of the midpoints of oat CENH3 ChIP-seq and input reads along the repeat clusters (*A*) CRA39 and (*B*) CRA82. The same CENH3 and canonical nucleosome-binding positions were observed in CENH3 ChIP and input samples. (*C*) Enrichment patterns of non–B-form DNA at CENH3 nucleosomes of the 21 chromosomes. (*D*) The distribution of CENH3 ChIP-seq and input reads along the non–B-DNA structures including DNA bending, slipped DNA, G-quadruplex, cruciform DNA, and Z-DNA with the up- and downstream regions. (*E*) Metaplot of CENH3 ChIP-seq reads along the R-loop peaks and with up- and downstream regions. (*F*) Snapshot of representative genomic regions showing the R-loop and CENH3-ChIP data. Line 1, unstranded (black) R-loops signal (gray); line 2, unstranded R-loop peaks; line 3, wR-loop signal (chartreuse); line 4, wR-loop peaks; line 5, cR-loop signal (magenta); line 6, cR-loop peaks, these signals are normalized to the genome-wide mean; line 7, CENH3 ChIP-seq reads; and line 8, CENH3 nucleosome positions. The triangle indicates the colocalization region of R-loops and CENH3 nucleosomes.

Although significant progress has been made in understanding centromere assembly, sequence characteristics, and its associated proteins, there are still many unanswered questions about centromere identity, maintenance, and evolution. The existence of neocentromeres implies that centromere function is largely independent of DNA sequences. One view holds that epigenetic regulation may be the major determinant of centromere specification. However, DNA structure is a critical element in determining its function in that a wide variety of non–B-form DNA structures affect gene expression, DNA synthesis and replication, and genome organization (12, 58). Evidence from an increasing number of studies indicates the important role of unusual DNA structures in the centromeres (21, 23). Thus, we would like to consider the role of DNA conformation in centromere organization and function. Using computational programs, we observed obvious enrichment of predicted Z-DNA, slipped DNA, and bent DNA at centromeres of the three oat subgenomes, and these predicted non–B-DNA structures were potentially formed in clustered transposons and CENH3-enriched regions (*SI Appendix*, Fig. S9*C*). Surprisingly, although some types of predicted non–B-form DNA were not enriched in oat centromeres compared with chromosome arms, such as G-quadruplexes, cruciform DNA, and H-DNA, these structures still occupy a large proportion of centromeres; we also observed the enrichment of CENH3 in the formation regions of G-quadruplexes and cruciform DNA (Fig. 5*C*). In Great apes and mouse, centromeric satellites have been predicted to have a low propensity to form cruciforms (59). However, cruciforms were associated with CENP-A enrichment in human and mouse when treated by permanganate (21). Thus, the potential roles of non–B-form DNA, which were not enriched in centromeres, cannot be discounted. Furthermore, cruciform structures were also detected at human and chicken neocentromeres (21–23), suggesting a role for centromere identity. One possible explanation is that CENH3 may be incorporated in loci where DNA sequences could adopt non–B-DNA conformations to initiate neocentromeres.

We also observed that the numbers of predicted non–B-DNA structures in the AA and DD subgenomes are greater than that in the CC subgenome (Fig. 2*C*). This is partly due to the fact that two of the three subgenomes are too homologous to distinguish from each other. These results suggest that the different non–B-DNA structures at centromeres may exhibit subgenome-biased distribution. Furthermore, predicted non–B-DNA structures varied when centromeres undergo rapid evolution from diploid to tetraploid to hexaploid oat (Fig. 6*E*). This may be explained by the fact that centromeric repeat variation occurred via deletion, insertion, and inversion or by mixing of different parental sequences during allopolyploid formation. These processes actively contribute to variability in retrotransposon nucleotide sequences, size, and number in oat centromeres (60, 61), and also result in the variation of non–B-DNA structures, affecting

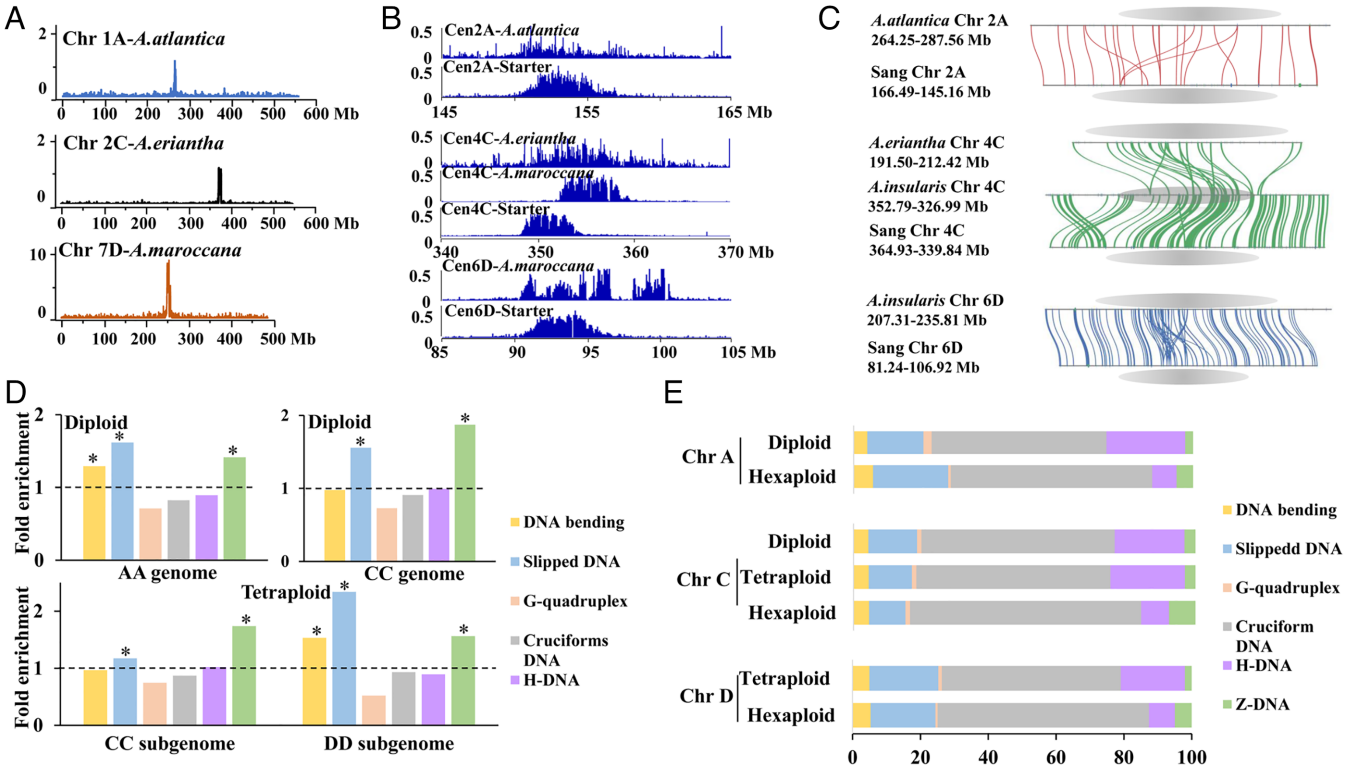

Fig. 6. Conformational changes of non–B-form DNA occurred during the evolution from diploid to tetraploid to hexaploid oat. (*A*) Genome-wide mapping of CENH3-ChIP-seq reads from diploid (*A. atlantica and A. eriantha*) and tetraploid (*A. maroccana*) oat to the AA, CC, and CCDD reference genomes (RPGC normalization in 5-kb bins). The X axis shows the positions on each chromosome. The Y axis represents the normalized sequence read count ratio between ChIP-seq and input on each position, which was calculated based on the ratio of reads per million/100. The different colored lines represent the read density in representative chromosomes (blue, Chr 1A of *A. atlantica*; black, Chr 2C of *A. eriantha*; and orange, Chr 7D of *A. maroccana*). (*B*) Genome-wide mapping of CENH3 ChIP-seq reads of diploid, tetraploid, and hexaploid oat to the AACCDD reference genome. The tracks show the distribution of CENH3 ChIP-seq read density of diploid, tetraploid, and hexaploid oat along Cen2A (145,000,000 to 165,000,000), Cen4C (340,000,000 to 370,000,000), and Cen6D (85,000,000 to 105,000,000) of the AACCDD reference genome using the Integrative Genomics Viewer tool, respectively. (*C*) Synteny alignment between centromeres of diploid (*A. atlantica* and *A. eriantha*), tetraploid (*A. insularis*), and hexaploid (Sang) oat chromosomes. The red lines represent the synteny gene pairs between *A. atlantica* and Sang in the AA subgenome. The green lines represent the synteny gene pairs between *A. eriantha*A. *insularis,* and Sang in the CC subgenome. The blue lines represent the synteny gene pairs between *A. insularis* and Sang in the DD subgenome. The centromeres are marked as gray ovals. (*D*) The fold enrichment of different types of centromeric non–B-form DNA relative to the whole subgenomes in diploid and tetraploid oat (permutation test; asterisks denote the observed value was >90% of the permutation value). The different colored lines represent different types of non–B-form DNA (yellow, DNA bending; blue, slipped DNA; orange, G-quadruplex; gray, cruciform DNA; purple, H-DNA; and green, Z-DNA). (*E*) The fraction of six types of non–B-form DNA in centromeric regions in diploid, tetraploid, and hexaploid oat.

centromere function. Despite the differences in sequence composition, the enrichment of predicted non–B-DNA structures was remarkably conserved in oat centromeres with different ploidy levels (Fig. 6*D*). Additionally, non–B-DNA structures are predicted to form in CENH3-binding regions, suggesting the potential roles of non–B-DNA structures in centromere function; DNA can adopt a number of alternative structures owing to the influence of special proteins. CENP-B was found to be involved in bending DNA (42). Yeast Scm3 and human HJURP are $CENH3^{CENPA}$ chaperones (62), which might read or stabilize non–B-DNA structures to deposit CENH3. Clearly, non–B-DNA structures alone do not necessarily attract CENH3.

R-loops have been studied among different species. The distribution of R-loops varies in species and subgenomes. Until recently, the analysis of whole-genome R-loop sequencing data was only described in three plant species, namely *Arabidopsis*, rice, and maize; their biological relevance in polyploids with large and complex genomes was still unknown. Here, we present genome-wide mapping of R-loops in hexaploid oat. Approximately 5% of the oat genome was covered by R-loop peaks; this proportion is much lower than that in *Arabidopsis*, rice, and maize. As expected, we observed high enrichment of R-loops in proximity to the centromeres and within pericentromeric heterochromatin, which are preferentially formed within repetitive sequences. These results are consistent with our previous work of maize R-loops, indicating that R-loops are common features of centromeres. Recent work showed that mitotic centromeric R-loops can recruit the ATR kinase to active aurora kinase B, which promotes accurate chromosome segregation (34). Further support for this model was recently provided by Moran et al. (63), and they extended this finding that R-loops produced at centromeric repetitive sequences localize the chromosome passenger complex to the inner centromere to maintain sister chromatid cohesion. We also found that R-loop formation regions were enriched in CENH3. These findings point toward the possibility that R-loops originating from centromeric repeats may serve as a placeholder for epigenetic signals and function.

Methods

Plant Materials. Oat seeds were germinated at room temperature for several days and then transplanted into soil and grown in the greenhouse with an average temperature of 20 °C under a 16-h light–8-h dark cycle.

ChIP and ChIP-seq. ChIP was conducted as previously described by Liu (64) with some modifications. A polyclonal antibody against oat CENH3 was used for ChIP experiments. Briefly, 3 to 5 g of young leaves was carefully ground in liquid nitrogen. Nuclei were isolated in lysis buffer (1 mM EDTA, 1 × cOmplete Mini EDTA-free

protease inhibitor cocktail, 10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 0.5% NP-40, 5 mM β-ME, and 0.1 mM PMSF). Pelleted nuclei were washed with lysis buffer and centrifuged at 1,200 g for 5 min at 4℃ twice. Pelleted chromatin was resuspended in MNase buffer (10% sucrose, 50 mM Tris-HCl, pH 7.5, 1 mM $CaCl_2$, 1 mM $MgCl_2$, and 0.1 mM PMSF) and digested with MNase (New England BioLabs, 0.5 U) at 37℃ for 20 min. The reaction was quenched with 0.5 M EDTA and centrifuged at 12,000 rpm for 10 min at 4℃. Fifty microliter protein A beads (Invitrogen 10002D) were added into the supernatant at 4℃ for 3 to 4 h while rotating to avoid nonspecific binding to the affinity beads. Then, input sample was taken from the supernatant, and CENH3 antibody was added into the remainder incubating at 4℃ overnight. Immunocomplexes were recovered by 50 μL protein A beads. Samples were washed three times each with low salt (50 mM NaCl, 50 mM Tris-HCl, 0.2 mM PMSF, 1 × complete Mini EDTA-free protease inhibitor cocktail, and 10 mM EDTA), medium salt (100 mM NaCl, 50 mM Tris-HCl, 0.2 mM PMSF, 1 × cOmplete Mini EDTA-free protease inhibitor cocktail, and 10 mM EDTA), and high salt (150 mM NaCl, 50 mM Tris-HCl, 0.2 mM PMSF, 1× cOmplete Mini EDTA-free protease inhibitor cocktail, and 10 mM EDTA). Elution was performed in elution buffer (50 mM NaCl, 20 mM Tris-HCl, 5 mM EDTA, and 2% SDS) at 65℃ for 55 min.

Sequencing libraries were generated using NadPrep DNA Library Preparation Kit for Illumina (1002101) and NadPrep UDI Adapter Kit Set C1 for Illumina (1003221) following the manufacturer's recommendations. Libraries were sequenced as 150-bp paired-end reads on the Illumina NovaSeq platform.

ssDRIP-seq and DRIP-qPCR. DNA:RNA immunoprecipitation assay was adapted from a previous study (35). Briefly, nuclei were isolated from 5 g of oat seedlings and then resuspended in 1.6 to 4 mL TE and lysed with SDS/proteinase K digestion at 37℃ overnight. The nucleic acid–containing extracts were mixed with phenol–chloroform, and genomic DNA was precipitated with salt–ethanol precipitation and then incubated at room temperature for 30 min. Precipitated DNA was gently spooled with pipette tips and washed with 70% ethanol, air-dried, and dissolved in 100 μL TE buffer. Fifty microliter DNA was sonicated on a Covaris system (10% Duty Factor, 200 cycles/burst, 140 peak incident power, and 60 s) to yield an average fragment size of 300 bp. Half of each sample was treated with RNase H (Takara 2151) at 37℃ overnight. DRIP was performed with S9.6 antibody (Kerafast ENH001). The DRIPed DNA was detected using fragment analysis with the NGS analysis mode. The ssDRIP-seq libraries were constructed from the DRIPed DNA with an average fragment size of ~200 bp using the Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences) according to the manufacturer's protocol. Library DNA was analyzed on a fragment analyzer and sequenced on an Illumina NovaSeq system using 2 × 150-bp sequencing. qPCR of the eluted DNA was performed on a Roche LightCycler 480 Instrument II using SYBR Green Mix (Roche 04887352001) and analyzed by using the comparative C_T method. The DNA:RNA hybrid enrichment was calculated based on the IP/input ratio.

DNA Probe Preparation and Labeling. The products of centromeric retrotransposons were amplified from oat using the aforementioned primers for CRA39, CRA82, CRA120, and CRA299. The PCR-amplified DNAs and ChIPed DNA were labeled with Alexa Fluor 488-dUTP (green) or Alexa Fluor 594-5-dUTP (red) as needed using the nick-translation method (65).

Production of Antibodies. Anti-CENH3 antibody was generated by immunizing rabbits with the synthetic peptide (RRPAATPAPGAPAQQRARK) based on the C-terminal region of oat CENH3.

Immunolocalization and FISH. Immunolocalization on root tips was performed essentially as described (35). For localization of S9.6, oat anthers were fixed overnight in ethanol:acetic acid (3:1, v/v) and then washed with 70% ethanol three times. Anthers at the pachytene stage were transferred to a tube containing enzyme solution and incubated at 37℃ for 20 min. Then, it was gently washed with 70% ethanol to dislodge the remaining unwanted cell types. Centrifuging 30 s in a mini-centrifuge removed nearly all the ethanol. Thirty microliters of 100% acetic acid were added, and the cell suspension was dropped onto a slide. The chromatin was cross-linked by exposure to UV light using a cross-linker (0.12 J). S9.6 antibody (1:100 dilution) was applied to the fixed sample overnight at 4℃. For colocalization of S9.6 antibody and CRA82, the primary and secondary antisera (goat anti-mouse antibody labeled by FITC green) were applied first, and then, FISH was performed with Texas red-5-dUTP-labeled CRA82. For CENH3 localization, oat root tips were obtained from newly germinated seeds and fixed by 4% formamide. Meiotic chromosomes were prepared for immunostaining. Procedures for the washing steps and mounting have been described previously (66). The root tip cells were treated with an equal amount of probe. In both cases, chromosomes were counterstained with DAPI. Images were acquired using confocal microscopy (Zeiss Cell Observer SD) and processed with Adobe Photoshop CS 6.0.

In Vitro Transcription of R-Loops and AFM Imaging. CRA39 (212-3878) and CRA82 (72-3853) derived from amplication of oat genomic DNA were cloned into the prokaryotic expression vector pET-30a according to standard protocols. Transcription reactions (50 μL total and 3 μg plasmid DNA) were carried out at 37℃ for 30 min in transcription buffer (1 × T7 RNA polymerase reaction buffer, 2 mM RRI, 1 mM DTT, and 0.5 mM rATP, rCTP, rUTP, and rGTP). Reaction was initiated by adding 40 mM T7 RNA polymerase. Enzyme was heat inactivated by incubating at 65℃ for 10 min. In addition, the random sequence (chr2A: 423505500-423507500) without R-loop formation was used as the negative control under the same conditions. RNase A was then added to digest soluble RNAs. The negative control was treated with both RNase A and RNase H and purified using phenol/chloroform extraction. The plasmid DNA was further digested with *Bam*HI and *Kpn*I to generate two fragments. For AFM sample preparation, DNA was diluted in ddH_2O and deposited on a freshly cleaved mica surface with aminopropyl silatrane (APS) for 15 min; the sample was washed with ddH_2O and dried under a gentle stream of nitrogen gas. Samples were measured using cantilevers (ScanAsyst-Air, Bruker) of tip radius 2 nm, resonance frequency of 70 kHz, and nominal force constant 0.4 N m^{-1}. All images were collected using the ScanAsyst mode on Bruker MultiMode 8 AFM with a nanoscope IIIa controller.

Dot Plot Analysis. The prepared AFM samples including the circular plasmids and the treated and untreated R-loop substrate with RNase A and RNase H were loaded onto the Hybond-N+ membrane (Amersham, RPN203B). After the membrane dried, nonspecific sites were blocked by soaking in 5% skimmed milk in Tris-buffered saline Tween-20 (TBST) at room temperature for 1 h. Then, it was incubated with anti-S9.6 antibody (100 μg/mL, 1:10,000 dilution) in 5% milk/TBST at 4 °C overnight, washed three times with TBST (3 × 5 min), and then incubated with secondary antibody (goat anti-mouse antibody conjugated to horseradish peroxidase, and 1:20,000 dilution in 5% milk/TBST, GE Healthcare, NA931) for 1.5 h at room temperature. Three washes with TBST (1 × 15 min and 2 × 5 min), and then once with Tris-buffered saline (5 min) were performed. Detection was performed using ECL reagent and exposed to X-ray film in the darkroom.

ChIP-seq and DRIP-seq Analysis. Quality control of raw reads was performed with FastQC software. Then, the adaptors and low-quality bases were removed with Trimmomatic v.0.36 (67). The trimmed reads were mapped to the reference genomes of *Avena sativa* (Sang) with BWA-MEM software using default parameters. Mapped reads with SAMtools view -F -20 were extracted for subsequent analysis, and duplicates were removed using SAMtools v.1.3.1 (44). For ssDRIP-seq, the total mapped reads were divided into forward (wR-loops, representing an R-loop formation containing an DNA:RNA hybrid on the Crick strand and ssDNA on the Watson strand) and reverse reads (cR-loops). MACS2 was used to detect peaks from the alignments (37). For visualization, the alignment BAM files were converted to normalized coverage files with 50-bp bins with deepTools (68). Snapshots of the data were constructed using the Integrative Genomics Viewer (69).

Identification and Characterization of Centromeric Repeats. A total of 1.2 million randomly selected whole-genome shotgun reads from hexaploid oat were analyzed using Web-based Galaxy RepeatExplorer software (https://repeatexplorer-elixir.cerit-sc.cz/galaxy/). Based on the number of reads in each repeat cluster, the genome proportion of each repeat cluster was determined. To identify the centromeric repeats, the CENH3 ChIP-seq and input reads were subsequently subjected to BLAST analysis against the database of these cluster repeats (-evalue 1e-8). The ratio was calculated based on the numbers of aligned reads from ChIP-seq and input data and represents the CENH3 enrichment level. The genome proportion (%) of each cluster was estimated based on the number of reads associated with each repeat cluster. Tandem Repeats Finder v5.02 (33) was used to identify minisatellites. The sequences of high-ratio clusters and minisatellites were subjected to BLAST analysis against the oat reference genome. The entire structure of centromeric repeat clusters is determined using LTR finder (70). Centromeric clusters were amplified and confirmed by sequencing. Then, the clones were used as probes for FISH. For

minisatellites, oligos modified at 5′-ends with carboxytetramethylrhodamine (TAMRA) were synthesized for use in FISH.

Detection of Dyad Symmetries and Prediction of DNA Secondary Structure in Genomic Regions. The European Molecular Biology Open Software Suite (EMBOSS) palindrome (71) was used to detect dyad symmetries with the following parameters: -minpallen 5 -maxpallen 100 -gaplimit 20 -overlap.

We calculated the total lengths of the palindromic regions containing the sequences of interest. The dyad density of a sequence was normalized by length and defined as the sum of the values.

RNAfold (v2.3.5) (72) was used to predict folding free energies of genomic regions with the following parameters: -noGU -noconv -noPS -paramFile¼dna_mathews2004.par -p -g.

Prediction of Z-DNA, Cruciform DNA, Slipped DNA, Triplex DNA, DNA Bending, and G-Quadruplex. Identification of different non–B-form DNA motifs was performed using the non–B-DNA motif search tool (nBMST; https://nonb-abcc.ncifcrf.gov/apps/nBMST/default/) (32) that can generate accurate predictions of non–B-form DNA motifs with default parameters. We focus on six non–B-form DNA motifs: inverted repeats, direct repeats, mirror repeats, alternating purine-pyrimidinetracts, G4 motifs, and A-phased repeats The whole-genome distribution of non–B-form DNA was plotted in 100-kb windows along 21 chromosomes.

Data, Materials, and Software Availability. The ChIP-seq and ssDRIP-seq data generated in this study have been submitted to the NGDC Genome Sequence Archive (GSA; https://bigd.big.ac.cn/gsa/) under accession numbers CRA007177 and CRA007190. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We thank Drs Howard W. Rines and Ronald L. Phillips (University of Minnesota, USA) for kindly providing the hexaploid oat seeds. We thank Yuanying Peng (Sichuan Agricultural University) for providing the diploid and tetraploid oat materials including *A. atlanticaA. eriantha*, and *A. maroccana*. We thank Nathan D. Han from The Edison Family Center for critical reading of the article and helpful comments. This work was supported by the National Natural Science Foundation of China (31991212 and 31920103006) and Taishan Scholarship (201812123).

Author affiliations: [a]Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; [b]University of Chinese Academy of Sciences, Beijing 100049, China; [c]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China; [d]National Laboratory for Condensed Matter Physics and Key Laboratory of Soft Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China; [e]Laboratory of Plant Chromosome Biology and Genomic Breeding, School of Life Sciences, Linyi University, Linyi 276000, China; [f]Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan 250000, China; and [g]Division of Biological Science, University of Missouri-Columbia, Columbia, MO 65211-7400

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

1. H. Hou *et al.*, Centromeres are dismantled by foundational meiotic proteins Spo11 and Rec8. *Nature* **591**, 671–676 (2021).
2. Q. Liu *et al.*, Emerging roles of centromeric RNAs in centromere formation and function. *Genes Genomics* **43**, 217–226 (2021).
3. J. Zhou *et al.*, Centromeres: From chromosome biology to biotechnology applications and synthetic genomes in plants. *Plant Biotechnol. J.* **20**, 2051–2063 (2022).
4. I. M. Cheeseman, The kinetochore. *Cold Spring Harb. Perspect. Biol.* **6**, a015826 (2014).
5. S. Henikoff, H. S. Malik, Centromeres: Selfish drivers. *Nature* **417**, 227 (2002).
6. H. Su *et al.*, Knl1 participates in spindle assembly checkpoint signaling in maize. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022357118 (2021).
7. S. Henikoff, K. Ahmad, H. S. Malik, The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
8. Y. Liu *et al.*, Phosphorylation of histone H3 by Haspin regulates chromosome alignment and segregation during mitosis in maize. *J. Exp. Bot.* **72**, 1046–1058 (2021).
9. Y. Liu *et al.*, Cohesion and centromere activity are required for phosphorylation of histone H3 in maize. *Plant J.* **92**, 1121–1131 (2017).
10. K. Ichikawa *et al.*, Centromere evolution and CpG methylation during vertebrate speciation. *Nat. Commun.* **8**, 1833–1843 (2017).
11. R. N. Douglas *et al.*, De novo centromere formation on chromosome fragments with an inactive centromere in maize (*Zea mays*). *Chromosome Res.* **29**, 313–325 (2021).
12. M. Crespi, F. Ariel, Non-B DNA structures emerging from plant genomes. *Trends Plant Sci.* **27**, 1360–1385 (2022).
13. A. T. Jonstrup *et al.*, Hairpin structures formed by alpha satellite DNA of human centromeres are cleaved by human topoisomerase IIalpha. *Nucleic Acids Res.* **36**, 6165–6174 (2008).
14. H. Li *et al.*, The cumulative formation of R-loop Interacts with histone modifications to shape cell reprogramming. *Int. J. Mol. Sci.* **23**, 1567–1581 (2022).
15. M. Ohno, T. Fukagawa, J. S. Lee, T. Ikemura, Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma* **111**, 201–213 (2002).
16. Y. Fang *et al.*, Characterization of functional relationships of R-loops with gene transcription and epigenetic modifications in rice. *Genome Res.* **29**, 1287–1297 (2019).
17. L. Gonzalo *et al.*, R-loops at microRNA encoding loci promote co-transcriptional processing of pri-miRNAs in plants. *Nat. Plants* **8**, 402–418 (2022).
18. G. Miglietta, M. Russo, G. Capranico, G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. *Nucleic Acids Res.* **48**, 11942–11957 (2020).
19. A. Marnef, G. Legube, R-loops as Janus-faced modulators of DNA repair. *Nat. Cell Biol.* **23**, 305–313 (2021).
20. E. Kejnovsky, V. Tokan, M. Lexa, Transposable elements and G-quadruplexes. *Chromosome Res.* **23**, 615–623 (2015).
21. S. Kasinathan, S. Henikoff, Non-B-form DNA is enriched at centromeres. *Mol. Biol. Evol.* **35**, 949–962 (2018).
22. V. S. P. Patchigolla, B. G. Mellone, Enrichment of non-B-form DNA at *D. melanogaster* centromeres. *Genome Biol. Evol.* **14**, evac054 (2022).
23. F. Chardon *et al.*, CENP-B-mediated DNA loops regulate activity and stability of human centromeres. *Mol. Cell* **5**, 1751–1767 (2022).
24. R. D. Wells, Non-B DNA conformations, mutagenesis and disease. *Trends Biochem. Sci.* **32**, 271–278 (2007).
25. B. R. Baum, *Oats: Wild and Cultivated, a Monograph of the Genus Avena L. (Poaceae)* (Minister of Supply and Services, Ottawa, 1977), pp. 1–463.
26. N. A. Tinker *et al.*, Genome analysis in *Avena sativa* reveals hidden breeding barriers and opportunities for oat improvement. *Commun. Biol.* **5**, 474–484 (2022).
27. P. J. Maughan *et al.*, Genomic insights from the first chromosome-scale assemblies of oat (*Avena spp.*) diploid species. *BMC Biol.* **17**, 92 (2019).
28. Y. Peng *et al.*, Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nat. Genet.* **54**, 1248–1258 (2022).
29. N. Kamal *et al.*, The mosaic oat genome gives insights into a uniquely healthy cereal crop. *Nature* **606**, 113–119 (2022).
30. X. Yang *et al.*, Amplification and adaptation of centromeric repeats in polyploid switchgrass species. *New Phytol.* **218**, 1645–1657 (2018).
31. H. Su *et al.*, Centromere satellite repeats have undergone rapid changes in polyploid wheat subgenomes. *Plant Cell* **31**, 2035–2051 (2019).
32. R. Z. Cer *et al.*, Non-B DB v2.0: A database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res.* **41**, D94–D100 (2013).
33. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
34. L. Kabeche, H. D. Nguyen, R. Buisson, L. Zou, A mitosis-specific and R loop-driven ATR pathway promotes faithful chromosome segregation. *Science* **359**, 108–114 (2018).
35. Y. Liu *et al.*, Genome-wide mapping reveals R-loops associated with centromeric repeats in maize. *Genome Res.* **31**, 1409–1418 (2021).
36. T. L. Bailey, STREME: Accurate and versatile sequence motif discovery. *Bioinformatics* **24**, 2834–2840 (2021).
37. Y. Zhang *et al.*, Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
38. Y. Liu *et al.*, Back-spliced RNA from retrotransposon binds to centromere and regulates centromeric chromatin loops in maize. *PLoS Biol.* **18**, e3000582 (2020).
39. H. Yan *et al.*, Intergenic locations of rice centromeric chromatin. *PLoS Biol.* **6**, 2563–2575 (2008).
40. H. Su *et al.*, Dynamic chromatin changes associated with de novo centromere formation in maize euchromatin. *Plant J.* **88**, 854–866 (2016).
41. A. Houben *et al.*, CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. *Chromosoma* **116**, 275–283 (2007).
42. D. Hasson *et al.*, The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat. Struct. Mol. Biol.* **20**, 687–695 (2013).
43. I. Ioshikhes, S. Hosid, B. F. Pugh, Variety of genomic DNA patterns for nucleosome positioning. *Genome Res.* **21**, 1863–1871 (2011).
44. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
45. O. Flores, M. Orozco, nucleR: A package for non-parametric nucleosome positioning. *Bioinformatics* **27**, 2149–2150 (2011).
46. N. A. Tinker *et al.*, Genome analysis in Avena sativa reveals hidden breeding barriers and opportunities for oat improvement. *Commun. Biol.* **5**, 474 (2022).
47. Q. Liu *et al.*, The repetitive DNA landscape in Avena (Poaceae): Chromosome and genome evolution defined by major repeat classes in whole-genome sequence reads. *BMC Plant Biol.* **19**, 226 (2019).
48. E. V. Ananiev, R. L. Phillips, H. W. Rines, Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13073–13078 (1998).
49. M. Naish *et al.*, The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science* **374**, eabi7489 (2021).
50. Z. Cheng *et al.*, Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14**, 1691–1704 (2002).
51. G. X. Wang *et al.*, ChIP-cloning analysis uncovers centromere-specific retrotransposons in *Brassica nigra* and reveals their rapid diversification in Brassica allotetraploids. *Chromosoma* **128**, 119–131 (2019).
52. Z. Gong *et al.*, Repeatless and repeat-based centromeres in potato: Implications for centromere evolution. *Plant Cell* **24**, 3559–3574 (2012).

53. W. H. Shang *et al.*, Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res.* **20**, 1219–1228 (2010).
54. S. Luo *et al.*, The cotton centromere contains a Ty3-*gypsy*-like LTR retroelement. *PLoS One* **7**, e35261 (2012).
55. Z. Yang *et al.*, Cotton D genome assemblies built with long-read data unveil mechanisms of centromere evolution and stress tolerance divergence. *BMC Biol.* **19**, 115–137 (2021).
56. T. Langdon *et al.*, Retrotransposon evolution in diverse plant genomes. *Genetics* **156**, 313–325 (2000).
57. T. K. Wolfgruber *et al.*, Maize centromere structure and evolution: Sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. *PLoS Genet.* **5**, e1000743 (2009).
58. Y. Du, X. Zhou, Targeting non-B-form DNA in living cells. *Chem. Rec.* **13**, 371–384 (2013).
59. F. Kouzine *et al.*, Permanganate/S1 nuclease footprinting reveals non-B DNA structures with regulatory potential across a Mammalian genome. *Cell Syst.* **4**, 344–356 (2017).
60. W. Singchat *et al.*, Population scale analysis of centromeric satellite DNA reveals highly dynamic evolutionary patterns and genomic organization in long-tailed and rhesus macaques. *Cells* **11**, 1953 (2022).
61. D. P. Melters *et al.*, Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, R10 (2013).
62. Z. Zhou *et al.*, Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature* **472**, 234–237 (2011).
63. E. C. Moran *et al.*, Mitotic R-loops direct Aurora B kinase to maintain centromeric cohesion. bioRxiv [Preprint] (2021), https://doi.org/10.1101/2021.01.14.426738 (accessed on 10 May 2022).
64. Y. Liu *et al.*, Sequential de novo centromere formation and inactivation on a chromosomal fragment in maize. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E1263–1271 (2015).
65. A. Kato, J. C. Lamb, J. A. Birchler, Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 13554–13559 (2004).
66. H. G. Yu *et al.*, Neocentromere-mediated chromosome movement in maize. *J. Cell Biol.* **139**, 831–840 (1997).
67. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
68. F. Ramirez *et al.*, deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).
69. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform.* **14**, 178–192 (2013).
70. Z. Xu, H. Wang, LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
71. P. Rice, I. Longden, A. Bleasby, EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
72. R. Lorenz *et al.*, ViennaRNAPackage 2.0. *Algorithms Mol Biol*. **6**, 1–14 (2011).