The *CentO* satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres

Tao Zhang[a,b,1], Paul B. Talbert[c,1], Wenli Zhang[a,1], Yufeng Wu[a], Zujun Yang[b], Jorja G. Henikoff[c], Steven Henikoff[c,2], and Jiming Jiang[a,2]

[a]Department of Horticulture, University of Wisconsin, Madison, WI 53706; [b]School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu 610054, Sichuan, China; and [c]Howard Hughes Medical Institute and Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Contributed by Steven Henikoff, October 17, 2013 (sent for review September 26, 2013)

Plant and animal centromeres comprise megabases of highly repeated satellite sequences, yet centromere function can be specified epigenetically on single-copy DNA by the presence of nucleosomes containing a centromere-specific variant of histone H3 (cenH3). We determined the positions of cenH3 nucleosomes in rice (*Oryza sativa*), which has centromeres composed of both the 155-bp *CentO* satellite repeat and single-copy non-*CentO* sequences. We find that cenH3 nucleosomes protect 90–100 bp of DNA from micrococcal nuclease digestion, sufficient for only a single wrap of DNA around the cenH3 nucleosome core. cenH3 nucleosomes are translationally phased with 155-bp periodicity on *CentO* repeats, but not on non-*CentO* sequences. *CentO* repeats have an ∼10-bp periodicity in WW dinucleotides and in micrococcal nuclease cleavage, providing evidence for rotational phasing of cenH3 nucleosomes on *CentO* and suggesting that satellites evolve for translational and rotational stabilization of centromeric nucleosomes.

CENP-A | nucleosome phasing | epigenetics | kinetochore

Centromeres, the chromosomal domains that attach to spindle microtubules to segregate eukaryotic chromosomes in mitosis and meiosis, are DNA elements bound by special nucleosomes that contain a centromere-specific variant of histone H3 (cenH3). In most plants and animals, cenH3 nucleosomes are found on centromeric DNA that comprises megabases of tandemly repeated "satellite" sequences. Despite this apparent preference for repetitive DNA, a fully functional centromere, called a neocentromere, can occasionally form by assembling cenH3 nucleosomes on a single-copy DNA sequence that was not previously part of a centromere, indicating that centromere specification is epigenetic in plants and animals (for reviews, see refs. 1–4).

The tandem arrays of highly repeated satellite sequences that compose most plant and animal centromeres can differ dramatically between closely related species (5), and even between different chromosomes (6–8), suggesting that satellite arrays undergo rapid evolution through expansions, contractions, gene conversions, and transpositions. Monomers of satellite repeats range in length from 5 bp in *Drosophila* to 1,419 bp in cattle although more than half of described monomers in 282 species have lengths between 100 and 200 bp, often regarded as approximately the length of nucleosomal DNA (6, 9). The cenH3 nucleosomes typically occupy only a portion of the satellite repeats, often in discontinuous blocks (7, 10–12), and the same or similar repeats often underlie flanking pericentromeric heterochromatin composed of conventional nucleosomes. Some of these repeats, for example African green monkey α-satellite DNA, have long been known to position conventional nucleosomes, resulting in arrays of regularly spaced nucleosomes, said to be translationally phased (13–15). Nucleosomes can occupy multiple alternative translational phases on the same satellite (16, 17). Translationally phased nucleosomal arrays have also been observed on satellites in cucumber and in several cereal species, where phasing varies among repeats and chromosomal regions (18, 19).

Recently deep-sequencing technology has been applied to centromeres treated with micrococcal nuclease (MNase), which preferentially digests linker DNA between nucleosomes, to determine the positioning of cenH3 nucleosomes on satellite repeats. In human cultured cells, substantial translational phasing of CENP-A, the human cenH3, was reported on α-satellite (20). In maize, a similar approach mapped CENH3 (the name used for plant cenH3s) on the 156-bp maize centromeric satellite *CentC* and on two retrotransposon-derived centromeric sequences, *CRM1* and *CRM2* (21). Evidence for translational phasing of CENH3 on *CentC* and *CRM1* was lacking, but 190-bp phasing was observed on *CRM2. CentC* was shown to have a strong periodicity of AA or TT dinucleotides about every 10 bp, which corresponds to one turn of the DNA double helix. This periodicity is thought to favor a particular orientation of the DNA toward the nucleosome core particle, based on DNA bendability, and is known as rotational phasing of nucleosomes (22–24).

Rice has centromeres characterized by the 155-bp satellite sequence *CentO*, which is related to maize *CentC* (25, 26). Although some rice centromeres have megabases of *CentO* satellites, other evolutionarily new centromeres have little *CentO,* so CENH3 nucleosomes are found on both *CentO* and non-*CentO* sequences (12). For example, *Cen8* is comprised of mostly non-*CentO* sequences and has a *CentO* array (*CentO_8*) that is

Significance

Centromeres are sites on chromosomes that mediate attachment to microtubules for chromosome segregation and often comprise tandemly repeated "satellite" sequences. The function of these repeats is unclear because centromeres can be formed on single-copy DNA by the presence of nucleosomes containing a centromere-specific variant of histone H3 (cenH3). Rice has centromeres composed of both the 155-bp *CentO* satellite repeat and single-copy non-*CentO* sequences. This study shows that rice cenH3 nucleosomes are regularly spaced with 155-bp periodicity on *CentO* repeats, but not on non-*CentO* sequences. *CentO* repeats have an ∼10-bp periodicity in dinucleotide pattern and in nuclease cleavage that suggests that *CentO* has evolved to minimize its bending energy on cenH3 nucleosomes and that centromeric satellites evolve for stabilization of cenH3 nucleosomes.

Author contributions: S.H. and J.J. designed research; W.Z. performed research; S.H. and J.J. contributed new reagents/analytic tools; T.Z., P.B.T., Y.W., Z.Y., J.G.H., S.H., and J.J. analyzed data; and T.Z., P.B.T., S.H., and J.J. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE50755).

See Commentary on page 19974.

[1]T.Z., P.B.T., and W.Z. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: steveh@fhcrc.org or jjiang1@wisc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1319548110/-/DCSupplemental.

spanned by a sequenced BAC (27). Centromeres like *Cen8* are thought to represent an intermediate stage in centromere evolution between rare neocentromeres that form on unique sequences and mature centromeres populated by megabase-sized arrays of satellites (7, 12). *Cen8* therefore presents an opportunity to compare the organization of CENH3 nucleosomes on *CentO* and non-*CentO* sequences. To that end, we used an antibody to rice CENH3 (27) to perform chromatin immunoprecipitation (ChIP) of CENH3 nucleosomes digested with MNase and sequenced the bound DNA (ChIP-Seq) to determine the positions of CENH3 nucleosomes on rice centromeres. We analyzed the sizes and positions of CENH3 nucleosomal DNA fragments on both *CentO* and non-*CentO* sequences to address the role of satellites in organizing centromeric chromatin and analyzed the sequence features of these fragments to look for evidence of nucleosome positioning signals.

Results

CENH3 Nucleosomes Are Well-Positioned on *Cen8*. To investigate the positions of CENH3 nucleosomes in rice centromeres, we performed anti-CENH3 ChIP using chromatin well-digested by MNase (4.0 Units) to ∼90% mononucleosome size to eliminate most linker DNA and gel-purified a mononucleosome band (∼80–230 bp). The immunoprecipitated DNA was sequenced using the Illumina Genome Analyzer II platform. We generated 39.6 million (M) 36-bp paired-end reads, of which 28.2 M mapped uniquely to the rice reference genome (TIGR7/IRGSP1). The distribution of fragment lengths (Fig. 1*A*) had a major peak at 127 bp, with four minor peaks ranging from 93 to 148 bp. For comparative analysis, we also generated 35.1 M 36-bp sequence reads from an MNase-digested rice genomic DNA library, and 273 M paired-end reads from an MNase-digested rice bulk mononucleosomal (input) DNA library, in which canonical H3-containing nucleosomes predominate. All sequencing data are available from GEO (accession no. GSE50755). The distribution of fragment lengths in the input mononucleosome library had a major peak at 147 bp and a minor peak at ∼134 bp (Fig. 1*A*). The relative distribution of fragment lengths for CENH3 nucleosomes and input mononucleosomes resembles that recently seen for native human cenH3 (CENP-A) nucleosomes and bulk nucleosomes (20) and indicates that cenH3 nucleosomes protect less DNA on average than canonical nucleosomes.

To identify the CENH3-binding domains in rice centromeres, we split each chromosome into 1-kb windows and plotted the $\log_2$ fold change between the anti-CENH3 ChIP-seq fragment count and the input mononucleosomal DNA fragment count within each window. The ChIP-seq fragment distribution patterns in the four best-sequenced rice centromeres (*Cen4*, *Cen5*, *Cen7*, and *Cen8*) were similar to those obtained previously using an anti-CENH3 ChIP-454 sequencing dataset (12). We observed many distinct peaks of mononucleosome size in several of the most CENH3-enriched subdomains in *Cen8* (Fig. 1*B*). One of these regions (12,964,000–12,967,000) contains the *CentO_8* satellite block and is highly enriched with CENH3 whereas another (13,430,000–13,434,000) contains a *Gypsy/DIRS1* transposon and is moderately enriched with CENH3. ChIP-seq fragment distribution at 1-bp resolution revealed well-positioned nucleosomal peaks identified by nucleR (28) in both regions, although with somewhat variable average spacing per well-positioned nucleosome (187 bp for region 12,964,000–12,967,000 or 154 bp for subregion 12,964,000–12,966,000; versus 307 bp for region 13,430,000–13,434,000, which has more poorly positioned nucleosomes). Similarly, well-positioned nucleosome peaks were observed throughout *Cen4*, *Cen5*, and *Cen7* (Fig. S1) using both fragment count and nucleR score, indicating that most CENH3 nucleosomes are well-positioned.

The CENH3 Nucleosome Cores Are Smaller than Canonical Nucleosome Cores. Fragments from MNase-digested chromatin can be used to map the boundary/cutting sites of individual nucleosomes. The MNase-resistant nucleosomal core can be visualized through counts of MNase cutting sites from the (+) and (−) strand-specific reads (29). Using this procedure, we aligned the nucleosome centers identified by nucleR from all well-positioned canonical nucleosomes from the input mononucleosomal DNA library and from CENH3 nucleosomes. We found that the distance between the (+) and (−) strand peaks was 147 bp for canonical nucleosomes, but the distance was ∼103–129 bp for CENH3 nucleosomes, depending on which of several closely-spaced peaks was measured (Fig. 2*A*). Because of the preferential

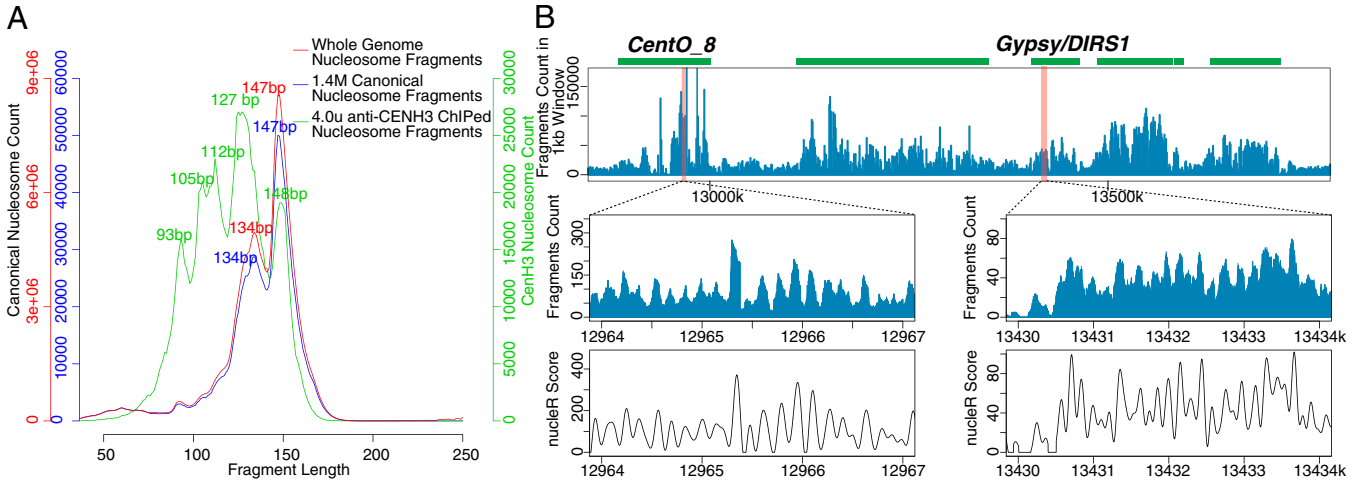

Fig. 1. CENH3 nucleosome positioning in rice *Cen8*. (*A*) Distribution of fragment lengths in size-selected 4.0 U MNase-treated mononucleosome libraries. The green line illustrates the distribution of all of the ∼1.4 M anti-CenH3 ChIP-seq fragments mapped to CENH3 subdomains in 12 rice centromeres. The blue line illustrates the distribution of 1.4 M randomly selected fragments from the input mononucleosomal DNA sequencing library. The red line illustrates the distribution of all mappable fragments (∼241 M) from the mononucleosomal DNA sequencing library. The *x* axis represents the length of fragments in base pairs. The *y* axis represents counts for each length. (*B*) Distribution of ChIP-seq fragments in *Cen8*. Vertical blue lines represent numbers of fragments within each 1-kb window. The horizontal green bars mark the CENH3-enriched subdomains. Two regions, *CentO_8* and a *Gypsy /DIRS1* transposon (pink boxes), are enlarged to illustrate CENH3 nucleosome positioning. NucleR scores of the two expanded regions are shown at the bottom.

digestion of AT-rich DNA by MNase (30), we used MNase-digested rice genomic DNA to calibrate the mapping of the ChIP-seq dataset. The result after calibration gave a single major peak for both (+) and (−) strands and an average distance of ∼112 bp (Fig. 2*B*).

The MNase cleavage peaks used to measure CENH3 nucleosomes are sensitive to the extent of digestion. To reduce the frequency of nucleosome-internal cleavages, which increases with MNase digestion, we performed two additional CENH3 ChIP-seq experiments using rice chromatin lightly digested with 0.5 and 2.5 Units of MNase, respectively. Application of 0.5 U of MNase resulted in underdigestion relative to typical ChIP-ready chromatin of mononucleosome size (Fig. 2*C*). Four libraries were constructed and paired-end sequenced, two from ChIPed and two from input DNA samples (Fig. S2*A*). Nearly identical patterns of positioning and spacing of the CENH3 nucleosomes were observed using the ChIP-seq data from the three independent ChIP experiments (Fig. S2*B*).We again observed that the average distance between the (+) and (−) strand peaks for CENH3 nucleosomes was shorter than that for canonical nucleosomes (Fig. S2*C*).

To confirm and better define the size of rice CENH3 nucleosomes, we made V-plots (31, 32) using ChIP-seq fragments associated with all well-positioned CENH3 nucleosomes aligned at their centers (Fig. 2 *D*–*F*). V-plots are based on paired-end read datasets generated by a modified Illumina library preparation protocol that permits efficient recovery of DNA fragments as small as ∼25 bp. On the V-plot, the *x* axis represents the distance from the fragment midpoint to the center of the CENH3 nucleosome, and the *y* axis represents the fragment length. The value on the *y* axis corresponding to the vertex of the V-plot represents the size of the CENH3 nucleosome core particle. The two V-plots using 0.5 U and 2.5 U of MNase both indicated that the protected region of CENH3 nucleosomes is ∼100 bp (Fig. 2 *D* and *E*) whereas the V-plot using 4.0 U of MNase had a vertex at ∼90 bp (Fig. 2*F*). This smaller size might indicate that the DNA on the CENH3 nucleosome has been more precisely trimmed back or that internal cleavage of CENH3 nucleosomes has reduced the size of the protected DNA in the 4.0-U MNase

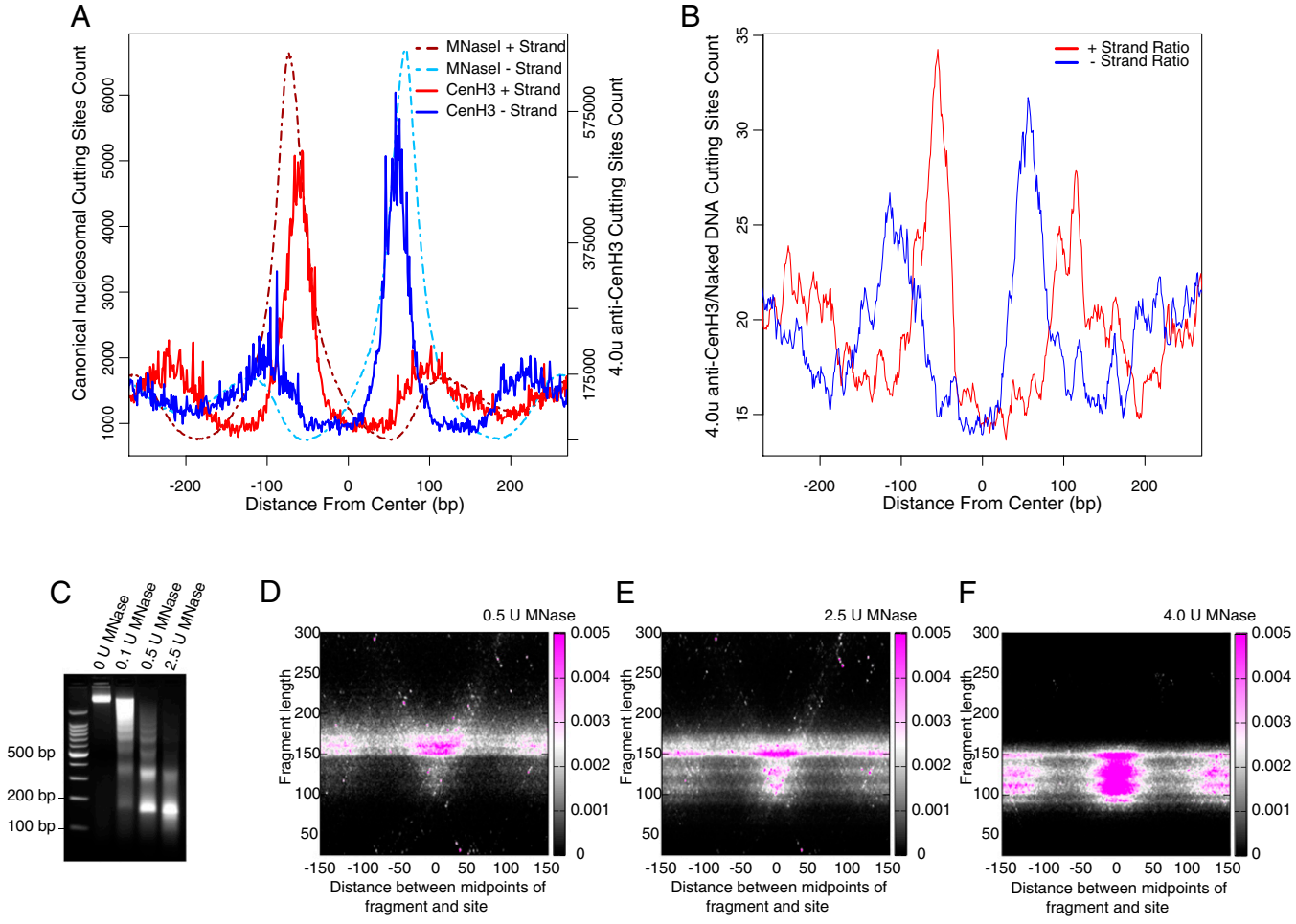

Fig. 2. Core sizes of CENH3 and canonical nucleosomes. (*A*) The nucleosome cores are anchored by mapping MNase cutting sites from (+) and (−) strands, (red and blue, respectively). The *x* axis represents the distance from the center. The *y* axis represents the count of MNase cutting sites. (*B*) CENH3 nucleosomes were mapped by using strand-specific ChIP-seq reads normalized by reads from MNase-digested rice genomic DNA. The *x* axis represents the distance from the center. The *y* axis represents the ratio of read count between ChIP-seq and MNase-digested rice genomic DNA. (*C*) Agarose gel electrophoresis of DNAs isolated from rice chromatin digested with different amounts of MNase. Chromatin samples digested with 0.5 U and 2.5 U MNase were used to make underdigested CENH3 ChIP-seq libraries. (*D*–*F*) V-plots of ChIP-seq fragments from libraries prepared using chromatin digested with (*D*) 0.5 U MNase, (*E*) 2.5 U MNase, or (*F*) 4.0 U MNase followed by size selection (∼80–230 bp). A 300-bp region spanning aligned CENH3 nucleosome centers from each library is shown. The heat maps were generated by quantifying the number of fragments of a given length (*y* axis) and given distance from the midpoint of fragments to centers (*x* axis). Color represents quantified fragment scores from the low (black) to high (pink).

digestion. In either case, the CENH3 nucleosome appears to wrap no more than 100 bp of DNA, sufficient for only a single wrap around the CENH3 nucleosome.

Intrinsic DNA Sequence Features Associated with CENH3 Nucleosomes. DNA sequences are intrinsically important for nucleosome positioning. Dinucleotides SS (G/C) and WW (A/T) were reported to be signals for nucleosome positions (22–24, 33). To examine whether rice CENH3 nucleosomes exhibit similar intrinsic DNA sequence features, we aligned the centers from 4,255 well-positioned CENH3 nucleosomes from the 4.0-U MNase digestion and analyzed the distribution of SS and WW dinucleotides within ±150 bp from the center. The center and its immediately flanking regions were enriched with SS dinucleotides. In contrast, WW dinucleotides were enriched in regions ±75 bp from the center (Fig. 3 *A* and *B*).

For comparison, we analyzed 265,487 well-positioned canonical nucleosomes from the 4.0-U library and observed a highly similar dinucleotide distribution pattern (Fig. 3*A*). However, the enriched WW dinucleotides were detected in regions ±91 bp away from the center, 16 bp further in each direction than for the CENH3 nucleosomes, consistent with our observations that CENH3 nucleosomes are smaller than conventional nucleosomes. These results agree with observations in humans that G/C-enriched regions favor nucleosome centering whereas flanking regions enriched for AA or TT dinucleotides possibly act as repelling elements to restrict the translational positioning of nucleosomes (34).

To investigate whether CENH3 nucleosomes on *CentO* also possess similar intrinsic DNA sequence features, we used ClustalW 2.0 (35) to align all *CentO* or *CentO*-like sequences of length ∼155 bp (available at ftp://ftp.plantbiology.msu.edu/pub/data/TIGR_Plant_Repeats/) and generated a 155-bp consensus *CentO* sequence (*SI Text*). This sequence was used to identify *CentO* repeats in the *CentO_8* satellite block from *Cen8*, using BAC a0038J12 (GenBank ID no. AY360388), which spans the *CentO_8* block of *Cen8* and is fully sequenced (27). We chose to use BAC a0038J12, which differs in sequence assembly from the *CentO_8* sequence of the International Rice Genome Sequencing Project, because much of the assembly of BAC a0038J12 has been verified through restriction digests and because the size of *CentO_8* on this assembly most closely matches the length of *CentO_8* on *Chromosome 8* as estimated from fiber-FISH (26, 27, 36). Seventy-one *CentO* units were identified within the 94,000- to 106,000-bp region of BAC a0038J12 (Fig. 4*A*).

We mapped all perfectly matched ChIP-seq fragments to this region, whether a fragment was mapped to a single or multiple sites in the BAC clone. The fragment count peaks in this region corresponded well with individual *CentO* monomers (Fig. 4*B*) and were similar to those located in other CENH3-enriched subdomains (Fig. 1 and Fig. S1). These results support the view that each *CentO* monomer is associated with a single CENH3 nucleosome. Nucleosome centers were aligned to examine the distribution of WW and SS dinucleotides within ±150 bp. We observed a similar SS/WW distribution pattern to that derived from single-copy centromeric sequences, superimposed on an ∼10-bp periodicity in the percentage of WW (Fig. 3*C*). A 10-bp periodicity of WW dinucleotides is thought to stabilize the rotational setting of the DNA double helix as it bends around the histone core (22, 37).

CENH3 Is Depleted from a 167-bp *CentO* Variant. BAC a0038J12 contains six copies of a 167-bp *CentO* variant that has a tandem duplication of 12 bp (CGAACGCACCCA). We determined the percentage of 11-mer sequences from *CentO* (both 155-bp and 167-bp variants) in fragments from anti-CENH3 ChIP, anti-H3K4me2 ChIP (38) and from input mononucleosome libraries. Then, we plotted the $\log_2$ fold change between the percentage of

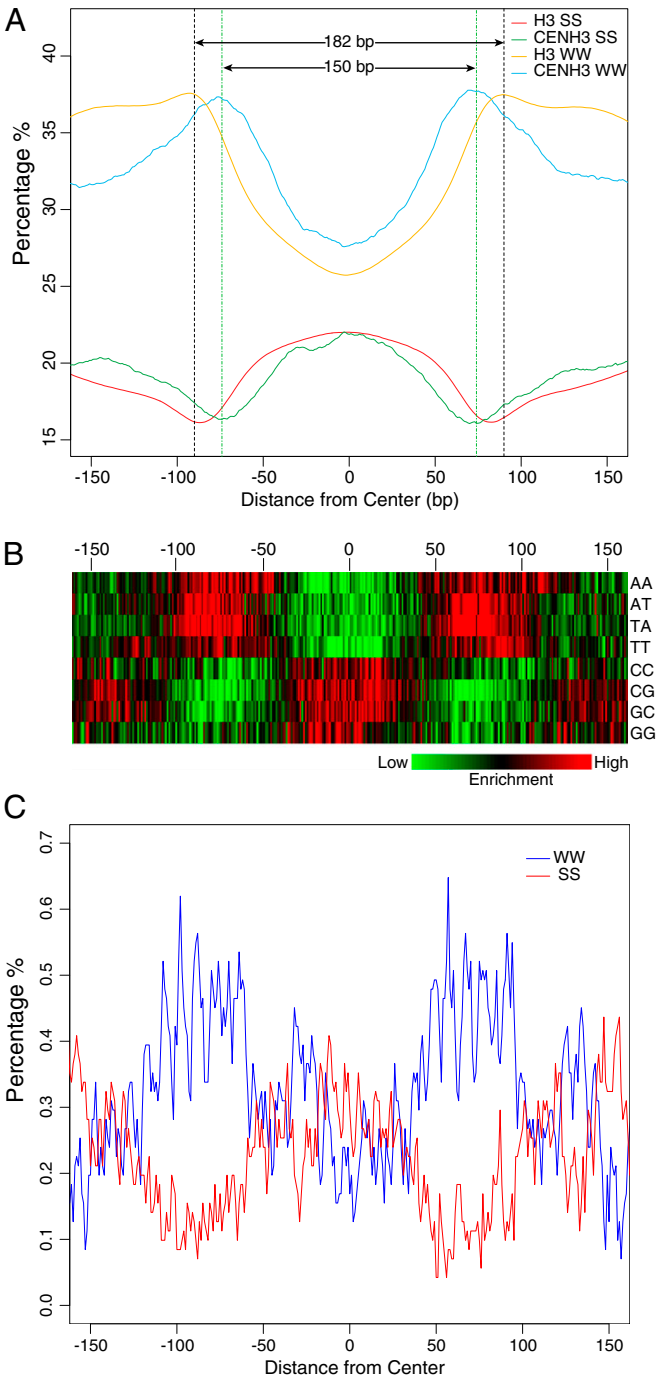

Fig. 3. Sequence features of CENH3 nucleosomal DNA. (*A*) Centers of positioned CENH3 and canonical nucleosomes are aligned, and sequence signals are displayed within ±150 bp from center. Each curve represents WW or SS dinucleotide frequency distributed in either CENH3 or canonical nucleosomes. The *x* axis represents the distance from the center; the *y* axis represents the frequency of SS/WW dinucleotides. (*B*) Heat map of eight different types of SS/WW dinucleotides ±150 bp from the CENH3 nucleosome center. (*C*) SS/WW dinculceotide frequency associated with *CentO*-related CENH3 nucleosomes, including 71 centers in *CentO_8* (94,000–106,000 bp in BAC a0038J12). The *x* axis represents the distance from the center; the *y* axis represents the frequency of SS/WW dinucleotides.

ChIPed fragments and input fragments for each 11-mer (Fig. 4 *C* and *D*). We found that, for the anti-CENH3 ChIP, the $\log_2$ fold change of two 11-mers (CACCCACGAAC and ACCCACGAA-

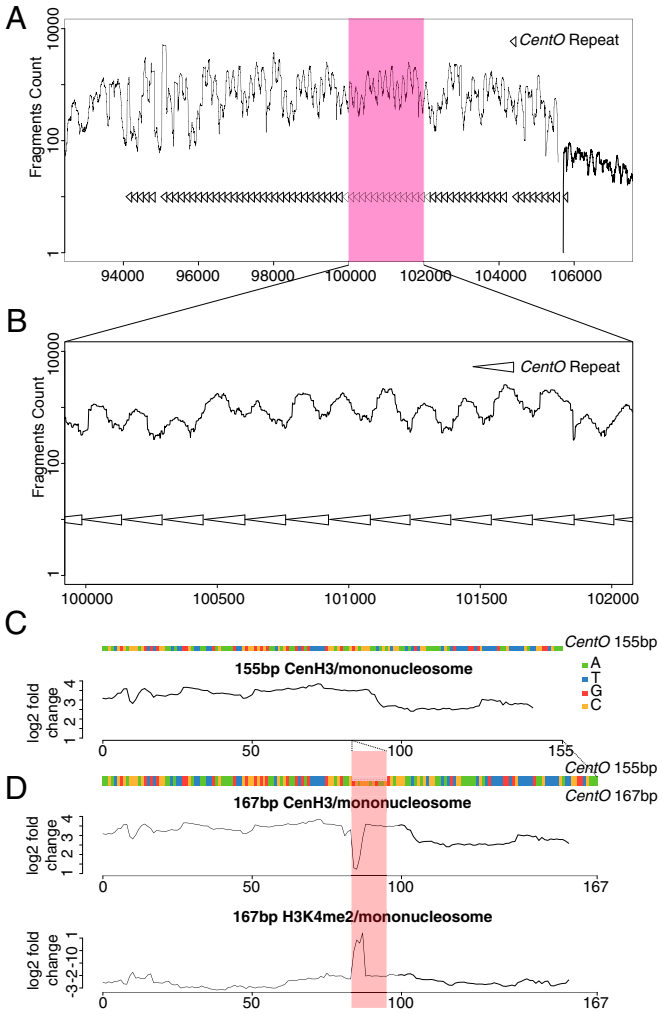

Fig. 4. Mapping of ChIP-seq fragments to *CentO* repeats in *CentO_8*. (*A*) Counts of ChIP-seq fragments (*y* axis) along the sequence of the *CentO_8* region (*x* axis). Locations of *CentO* repeats are indicated by arrowheads. (*B*) Enlargement showing one peak of fragment counts per *CentO*. (*C* and *D*) CENH3-nucleosomes on 155-bp and 167-bp *CentO* variants. (*C*) Enrichment in proportion of *CentO*-derived 11-mers between anti-CENH3 ChIPed and input mononucleosome fragments. The *x* axis represents the center position of 11-mers along the *CentO* consensus sequence (shown in color: A, green; T, blue; G, red; C, orange). The *y* axis represents the $\log_2$-fold change between ChIPed and mononucleosome libraries in the proportion of the 11-mers present in fragments. (*D*) The $\log_2$ fold change in proportion of 11-mers between anti-CENH3 ChIPed and mononucleosome fragments (*Upper*), or between anti-H3K4me2 ChIPed and mononucleosome fragments (*Lower*) along the 167-bp *CentO* sequence (shown together with the 155-bp *CentO* consensus for comparison).

CG) that are unique to the duplication junction of the 167-bp *CentO* sequence were strongly reduced compared with other flanking 11-mers present in both the 155-bp and 167-bp variants. Conversely, we found that these 11-mers were enriched in the H3K4me2 library over input. These results suggest that CENH3 nucleosomes prefer the 155-bp *CentO* variant and are depleted from the 167-bp *CentO* variant, which is preferentially occupied by H3 nucleosomes dimethylated on lysine 4.

CENH3 Nucleosomes Are Translationally Phased with 155-bp Periodicity on *CentO*. We generated phasograms (33, 34) by plotting the frequency of occurrence of each distance between the fragment midpoints of two ChIP-seq fragments within a window of the reference genome (Fig. S3*A*).

The phasogram of input mononucleosomal DNA fragments indicated that the average distance between two adjacent nucleosomes is 192 bp ($R^2 = 0.9997$, P value $= 2.4 \times 10^{-8}$) in the rice genome, and a virtually identical spacing was found for mononucleosomes on *Chromosome 8*, exclusive of CENH3-enriched regions (Fig. S3 *B* and *C*). In contrast, phasograms of CENH3 nucleosomes from the CENH3-enriched regions of *Cen8* that lack *CentO* sequences showed no clear global periodicity (Fig. 5*A*) although we cannot exclude the possibility that small sets of nucleosomes might be regularly spaced.

Sequence identity among *CentO* monomers ranges from 76% to 100% (36) so many *CentO* fragments are unique. To examine the internucleosomal spacing of CENH3 nucleosomes on *CentO*, unique fragments from all CENH3 nucleosomes were mapped onto *CentO_8*, and phasograms were generated. This analysis revealed that the average distance from the middle of one CENH3-protected particle to the middle of the adjacent CENH3-protected particle is ~155 bp ($R^2 > 0.9999$, P value $= 8.9 \times 10^{-10}$) (Fig. 5*A*). To determine whether this 155-bp periodicity of CENH3 peaks seen for *CentO_8* extends to *CentO* genome-wide, the midpoints of all unique fragments were mapped onto four tandem copies of the *CentO* consensus. A clear peak was seen on each *CentO* in the midpoint counts (Fig. S4*A*). We conclude that *CentO* imposes a preferred translational position with 155-bp periodic phasing on CENH3 nucleosomes.

Subnucleosome-Sized Particles Map Between Phased CENH3 Nucleosomes on *CentO_8*. We noticed that smaller subpeaks could be discerned between successive 155-bp peaks in *CentO_8* phasograms (Fig. 5*A*). Arbitrary separation of *CentO* fragments into >127-bp and <127-bp size classes revealed that these interstitial subpeaks are seen only in the phasogram representing the larger size fragments (Fig. 5*B*) whereas the phasogram of smaller fragments showed only the major 155-bp periodicity (Fig. 5*C*). Consistent with the double periodicity of the phasogram of the longer fragments, midpoint-counts of the >127-bp fragments mapped to the *CentO* consensus showed two peaks per *CentO* repeat, and both were offset from the position of the dominant peak seen in the <127-bp fragments (Fig. S4 *B*–*D*). The absence of the dominant midpoint-count peak in the longer >127-bp fragments suggests that the midpoints of longer fragments have shifted from the CENH3 nucleosome centers because linker DNA flanking these nucleosomes is also being protected in the >127-bp fragments.

We hypothesized that one or more labile particles that can protect linker DNA from MNase are located between the CENH3 nucleosomes and are pulled down by the anti-CENH3 antibody with CENH3 nucleosomes when they are present on the same DNA fragment, as has been observed in budding yeast (32). These longer protected fragments would include DNA protected by particles on the "left" or "right" side of an adjacent CENH3 nucleosome, and the phasogram peaks and subpeaks would then originate from plotting the distances between the midpoints of these two classes of fragments (see the model in Fig. S5).

To attempt to detect such an internucleosomal particle(s), which might represent other kinetochore proteins, we plotted the distributions of lengths of fragments in the 0.5-U and 2.5-U libraries that crossed nucleosome peak centers in *CentO_8* (Fig. 6*A*) and that crossed the point halfway between the nucleosome centers (peak + 78 bp) (Fig. 6*B*). A peak of ~35-bp fragments in the counts of peak + 78 bp fragments suggests that a particle pulled down with CENH3, perhaps through protein–protein contacts, protects DNA fragments of this size. The particle was enriched in the input mononucleosome libraries, indicating that the majority of it dissociated from CENH3 nucleosomes in the

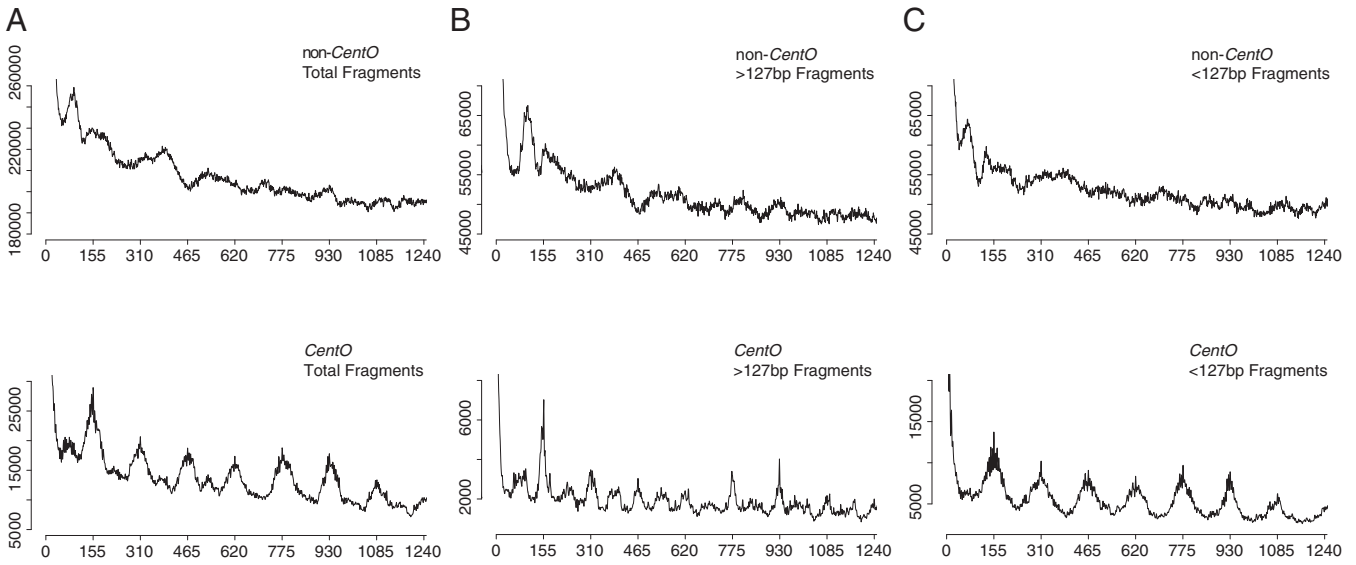

Fig. 5. CENH3 nucleosome phasing in rice *Cen8*. Phasograms were generated using ChIP-seq fragments from *Cen8* for non-*CentO* (*Upper*) and *CentO_8* (*Lower*) sequences. The *x* axes show the range of recorded phases. The *y* axes represent frequencies of distance between the midpoints of two fragments using (*A*) the total fragments, (*B*) >127-bp fragments, or (*C*) <127-bp fragments.

MNase digestion, consistent with its being a labile particle between CENH3 nucleosomes. When we made a similar fragment-length plot for fragments at the peak + 78 bp position for all *CentO*, a peak at ∼35 bp was not evident, although fragments in this general size range were strongly enriched in the input libraries relative to anti-CENH3 libraries (Fig. S6). Therefore, the labile particle that protects ∼35 bp between successive CENH3 nucleosomes may be specific to *CentO_8*.

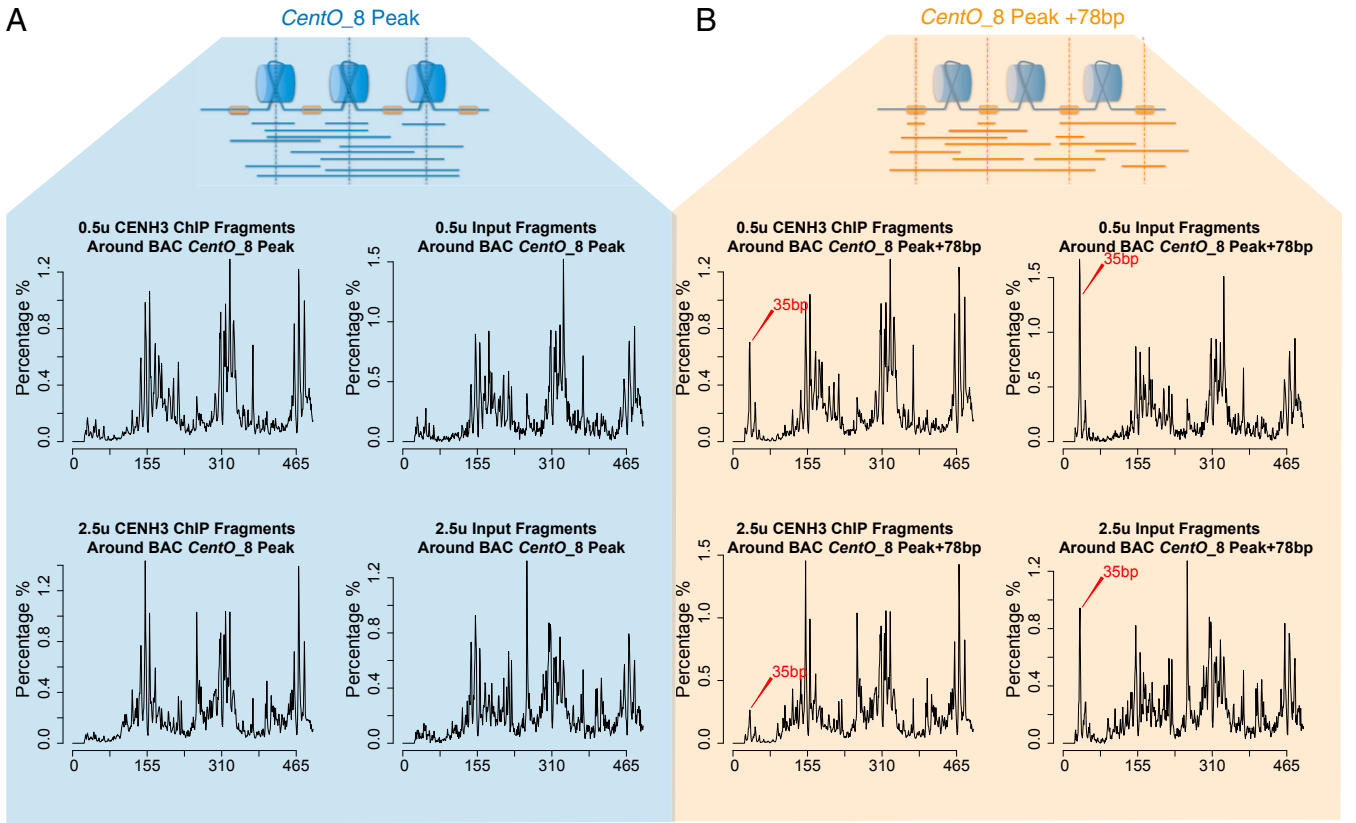

Fig. 6. Fragment-length distribution of *CentO* fragments from *CentO_8*. (*A*) Fragments that pass through the nucleosome peak or (*B*) the halfway point between nucleosome peaks (peak + 78 bp). The proportion of 35-bp fragments in *B* is decreased in anti-CENH3 libraries (*Left*) compared with input libraries (*Right*) and also decreases with increasing MNase.

***CentO* Nucleosomes Have an ~10-bp Periodicity of MNase Cleavage.** To further address the size and position of protected particles on *CentO*, we made V-plots using all CENH3-enriched *CentOs* aligned on the *CentO* consensus, using the 0.5-U or 2.5-U libraries, which were not size-selected. Surprisingly, although a 155-bp periodicity was observed in the V-plots, the dominant pattern was a 10-bp periodicity of MNase cleavage, yielding fragments as small as 20 bp, the limit of recovery (Fig. 7). We infer that the small fragments observed are due to internal cleavage within CENH3 nucleosomes. This 10-bp periodicity is also evident in the *CentO* fragment length distribution plots (Fig. 6 and Fig. S6). Such a strong 10-bp periodicity is remarkable, given that the *CentO* V-plot cleavage pattern is very different from the patterns seen under the same light MNase-digestion conditions for V-plots of CENH3 protection that also include non-*CentO* sequences (Fig. 2 *D–F*). The pattern of cleavage also differs from MNase digestion of naked *CentO* sequences (Fig. S7). The regular internal cleavage makes it impossible to use V-plots to infer the sizes of protected particles on *CentO*. Internal nucleosome cleavage by MNase with 10-bp periodicity has been attributed to the cleavage of accessible nucleotides that lie further from the surface of the histone core at each helical turn as DNA wraps around the core (39). The regular cleavage pattern observed therefore most likely reflects rotational phasing of CENH3 nucleosomes on *CentO*.

Discussion

Size of Protected DNA on CENH3 Nucleosomes. Rice centromeres provide the opportunity to examine CENH3 nucleosomes over both *CentO* satellite repeats and non-*CentO* centromeric sequences. Our ChIP-Seq fragments from *Cen8* chromatin revealed well-positioned nucleosomes over both *CentO* and non-*CentO* sequences, confirming and refining previous mapping of CENH3-enriched regions (12). Aligning these well-positioned nucleosomes allowed us to map the left and right cut sites around CENH3 nucleosomes, which indicated a protected region of about 112 bp. A precise size estimate was made by constructing V-plots, in which the minimum protected size of CENH3 nucleosomes was found to be ~100 bp even in lightly digested chromatin. This minimum size is similar to that recently reported for ChIPed nucleosomes containing the human cenH3, CENP-A (20), and budding yeast cenH3, Cse4 (32). The protected sizes of DNA on all of these cenH3 nucleosomes are smaller than the measured sizes (125–150 bp) (40–42) of DNA protected by CENP-A octameric nucleosomes, including the 147 bp protected by octameric CENP-A nucleosomes assembled on human centromeric α-satellite (43). The composition of cenH3 nucleosomes has been controversial, with the proposal of tetrameric, hexameric, and octameric models of the arrangement of subunits (44). Our results do not directly address the composition of rice CENH3 nucleosomes but indicate that these nucleosomes protect only enough DNA for a single DNA wrap around a nucleosome core and protect less DNA than the measured sizes of DNA protected by octameric CENP-A nucleosomes.

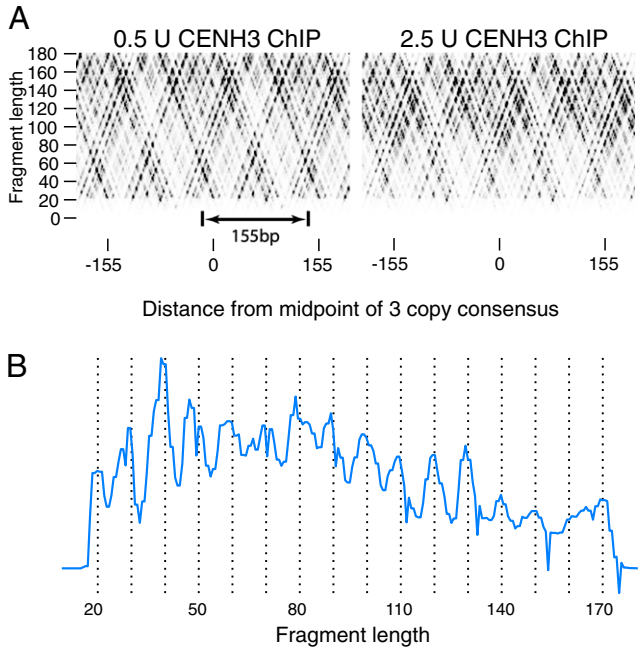

Fig. 7. Ten-base-pair periodicity of MNase cleavage on *CentO*. (*A*) The distances from the midpoints of fragments from the 0.5 U and 2.5 U libraries to the center of three copies of the *CentO* consensus sequence (*x* axis) were mapped against the lengths of the fragments (*y* axis) to generate V-plots. (*B*) Counts of fragment lengths show a 10-bp periodicity of MNase cleavage on *CentO*.

Translational and Rotational Phasing on *CentO* Satellite. In the *CentO_8* satellite block of *Cen8*, we observed that a single peak of CENH3 ChIP fragments was found on each 155-bp copy of *CentO*. We used phasograms to determine that the bulk of CENH3 nucleosomes were spaced with a periodicity of ~155 bp on *CentO_8*, but not on non-*CentO* sequences, suggesting a fairly precise translational phasing on *CentO* repeats. This interpretation was supported by mapping fragment midpoints genome-wide onto the *CentO* consensus, which revealed a single major peak of midpoint-counts on *CentO*, implying that there is a preferred translational position of CENH3 on *CentO* throughout the genome.

In contrast, longer fragments showed a double periodicity of phasogram peaks per *CentO* repeat, and two peaks of midpoint-counts per *CentO,* both offset from the peak of midpoint-counts in the smaller fragments. This result implies the presence of CENH3 nucleosomes that protect the linker DNA on the left or right sides of the nucleosome DNA and give rise to two sets of fragments that result in the double periodicity per *CentO* of the midpoint-count peaks and the phasogram peaks. Such extended protection might occur if MNase sometimes cuts on only one side of a DNA-binding protein(s) between CENH3 nucleosomes, which thereby extends protection from MNase to the right or left. In support of this model, we found evidence for a particle that protects ~35 bp of DNA between CENH3 nucleosomes in *CentO_8*, although it was found preferentially in the input libraries. This 35-bp particle may have been recovered in anti-CENH3 libraries through protein–protein contacts with CENH3 nucleosomes even though it was not on the same DNA fragment as a CENH3 nucleosome. Future investigation of the locations of other DNA-binding kinetochore proteins may help to test this model.

Our results are paralleled by recent data from human cell culture that indicate that CENP-A nucleosomes are largely translationally phased on centromeric α-satellite and that shorter and medium-length fragments (100–119 bp and 120–139 bp, respectively) showed better phasing than longer fragments (140–160 bp) (20), consistent with our observations on how fragment length relates to translational phasing. On maize *CentC*, a 156-bp periodicity in MNase cleavage was observed although the same periodicity in the cleavage of naked *CentC* DNA obscured whether the cleavage pattern reflected translational phasing (21). In the maize study, only fragments 145–175 bp were selected so phasing similar to that seen in shorter fragments in human and rice centromeres might have been missed. In maize, *CentC* showed an ~10-bp periodicity in AA/TT dinucleotides (21), similar to that observed for WW dinucleotides in *CentO*. Such periodicity of WW dinucleotides for each 10.4-bp turn of the DNA double helix is believed to facilitate bending around nucleosomes and

favor rotational phasing (22, 23, 45). In rice, we observed not only a 10-bp periodicity in WW dinucleotides in the *CentO* sequence, but also a 10-bp periodicity in the MNase cleavage pattern of CENH3 nucleosomes on *CentO* that was absent from the cleavage pattern when non-*CentO* nucleosomes were included. Although MNase cuts preferentially within the linker DNA between nucleosomes, it also cuts internally in nucleosomes with an ∼10-bp periodicity at sites that are accessible because they lie further from the surface of the histone core (39). We conclude that the cleavage pattern of MNase on *CentO* (Fig. 7) provides in vivo evidence for rotational phasing of CENH3 on *CentO*.

Implications for Evolution of Centromeric Satellites. Satellite sequences are known to undergo rapid expansions or contractions (7, 8, 36), which has been attributed to their "selfish" competition for preferential transmission in female meiosis, or "centromere drive" (2, 46). However, the reasons for the success or failure of any particular satellite in this competition have remained obscure. Nor has it been clear why satellites populate centromeres or how new satellites arise. Our evidence that *CentO* confers translational and rotational phasing on rice CENH3 nucleosomes suggests that regular positioning of cenH3 nucleosomes is advantageous for centromere formation. The 10-bp periodicity of WW dinucleotides is believed to stabilize nucleosomes through reducing the deformation energy required to wrap DNA around them (37), and the strongly positioning *Widom 601* sequence derived from α-satellite stabilizes conventional nucleosomes relative to other DNA sequences, such as 5S rDNA (47). These observations suggest that satellites may evolve to stabilize cenH3 nucleosomes against the pulling forces they undergo during chromosome segregation. Stabilizing cenH3 nucleosomes in a regular structure on satellite repeats may help to prevent the loss of cenH3 nucleosomes that are under tension, and to facilitate formation of the kinetochore.

Stabilization of cenH3 nucleosomes can explain how new satellites arise and populate centromeres. Tandem duplication of a sequence of any length with a selective advantage for cenH3 stabilization can potentially lead to its expansion into a satellite array through unequal crossover or other modes of recombination (48). The proposed transition of evolutionarily new centromeres that form initially as neocentromeres on unique sequences into mature satellite-based centromeres (3, 49) could then be effected either through transposition of existing satellites from another location to the neocentromere, or through duplication of endogenous neocentromere sequences (49, 50) that stabilize cenH3 nucleosomes, followed by expansion and fixation of the stabilizing repeat through recombination and centromere drive. Although a new satellite sequence would usually be at least long enough to wrap a cenH3 nucleosome, even smaller sequences could contribute to the ∼10-bp periodicity of WW dinucleotides that stabilizes rotational positioning of cenH3 nucleosomes, as has been suggested for the 20-bp *CentAs* satellite repeats of *Astragalus sinicus* (51). Satellites longer than the length of two or three nucleosomes are rare (9), perhaps because longer sequences are unlikely to contribute to stabilization equally throughout their length, putting them at a disadvantage to derivative expansions of their own most stabilizing subsequences.

The total size of CENH3 domains in grass species correlates with genome size, and chromosomes of different sizes in the same species tend to have CENH3 domains of a similar size, even if they have different sizes of satellite arrays. This relationship may be explained by a limiting component (52), which may be CENH3. If the expansion of a stabilizing satellite directs CENH3 to a particular centromere, it might deplete CENH3 from other centromeres. Redistributing CENH3 through compensatory expansions or contractions of satellites on the same or other chromosomes or destabilizing selfish satellite interactions through adaptation of kinetochore proteins (53–56) may help to ensure that a functional centromere is present on each chromosome. Recurrent cycles of expansions of stabilizing satellites seeking a genetic advantage in female meiosis followed by restoration of epigenetic centromere specification by adaptation of kinetochore proteins may explain the rapid evolution of satellites and kinetochore proteins (57) and may resolve the paradox of the common occurrence of satellites at centromeres, which nevertheless are epigenetically specified.

Materials and Methods

Materials. Rice cultivar "Nipponbare" seeds were germinated at room temperature for 3 d. Germinated seeds were then sowed in soil to continue to grow in the greenhouse. Two-week-old seedlings were collected for nuclei and DNA isolation.

The antibody to the N terminus of rice CENH3 has been previously described (27). The antibody recognizes an epitope on CENH3 that appears to be accessible in all chromosomal contexts, not only on *CentO* and other rice sequences, but in CENH3s of all grass species that have been tested (for example, ref. 52).

MNase Digestion, ChIP, and ChIP-seq. The collected seedlings were ground into fine powder in liquid nitrogen. The ChIP procedure, including nuclei isolation, MNase digestion, immunoprecipitation using anti-cenH3 antibodies, and recovery of ChIPed DNA for development of ChIP-seq libraries, was carried out by following the published protocols (58), except that purified rice chromatins were differentially digested using 0.5 U, 2.5 U, and 4.0 U of MNase (Sigma; N5386-200U), respectively. ChIPed DNA and input DNA corresponding to 0.5 U and 2.5 U MNase digested chromatins were used for customized ChIP-seq library development following the published protocols (31), which can recover insert sizes down to 20 bp.

Collection of mononucleosome-sized DNA for sequencing was performed as described previously (29). Rice chromatin was digested by MNase into ∼90% mononucleosomes plus 10% dinucleosomes. Mononucleosome-sized DNA fragments were gel purified for library preparation. Nipponbare genomic DNA was also digested by MNase to the same relative extent as DNA from the chromatin digestion, and 100–200 bp DNA were cut from 2% agarose gel and purified for library construction. All libraries were sequenced in the paired mode using the Illumina platform. All sequencing data are available from GEO (accession no. GSE50755).

Data Analysis. DNA fragments were mapped to the rice genome (TIGR7/IRGSP1) (59, 60), with Bowtie (61) reporting nonmismatches and uniquely mapped sites. The centers of CENH3 and canonical nucleosomes were identified by nucleR (28). V-plots were generated by following our previous study (31, 32). Phasograms were based on calculating the frequency of nucleosome distance between midpoints of all fragments in specific windows. The counting result was plotted as a phasogram, a waveform-like histogram, which represents frequency of distance between all nucleosomal DNA fragments in a specific window. The peaks of phasograms reflect the periodicity or phasing distance between the nearest-neighbor nucleosome, next-nearest neighbors, etc. We used the phasogram peaks to calculate the nucleosome spacing by applying the linear-fitting model from R. The K-mer (11-mer) count was calculated by applying JELLYFISH (62). All ChIP-seq fragments derived from the 4.0 U MNase digestion were mapped to four tandem copies of the *CentO* consensus using Novoalign. We randomly kept one mapping site, if one fragment was mapped to multiple sites. Data processing and analysis were done using Perl and R.

ACKNOWLEDGMENTS. We thank the Dale Bumpers National Rice Research Center for providing the Nipponbare seeds. This research was supported by National Science Foundation Grants DBI-0603927 and DBI-0923640 (to J.J.) and by the Howard Hughes Medical Institute.

1. Plohl M, Luchetti A, Mestrović N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409(1-2):72–82.
2. Malik HS, Henikoff S (2009) Major evolutionary transitions in centromere complexity. *Cell* 138(6):1067–1082.
3. Kalitsis P, Choo KH (2012) The evolutionary life cycle of the resilient centromere. *Chromosoma* 121(4):327–340.
4. Birchler JA, Han F (2013) Centromere epigenetics in plants. *J Genet Genomics* 40(5):201–204.
5. Lee HR, et al. (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proc Natl Acad Sci USA* 102(33):11793–11798.

6. Lohe AR, Hilliker AJ, Roberts PA (1993) Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics* 134(4):1149–1174.
7. Gong Z, et al. (2012) Repeatless and repeat-based centromeres in potato: Implications for centromere evolution. *Plant Cell* 24(9):3559–3574.
8. Rudd MK, Wray GA, Willard HF (2006) The evolutionary dynamics of alpha-satellite. *Genome Res* 16(1):88–96.
9. Melters DP, et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* 14(1):R10.
10. Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. *Dev Cell* 2(3):319–330.
11. Zhong CX, et al. (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 14(11):2825–2836.
12. Yan H, et al. (2008) Intergenic locations of rice centromeric chromatin. *PLoS Biol* 6(11): e286.
13. Musich PR, Maio JJ, Brown FL (1977) Subunit structure of chromatin and the organization of eukaryotic highly repetitive DNA: Indications of a phase relation between restriction sites and chromatin subunits in African green monkey and calf nuclei. *J Mol Biol* 117(3):657–677.
14. Maio JJ, Brown FL, Musich PR (1977) Subunit structure of chromatin and the organization of eukaryotic highly repetitive DNA: Recurrent periodicities and models for the evolutionary origins of repetitive DNA. *J Mol Biol* 117(3):637–655.
15. Musich PR, Brown FL, Maio JJ (1982) Nucleosome phasing and micrococcal nuclease cleavage of African green monkey component alpha DNA. *Proc Natl Acad Sci USA* 79(1):118–122.
16. Zhang XY, Fittler F, Hörz W (1983) Eight different highly specific nucleosome phases on alpha-satellite DNA in the African green monkey. *Nucleic Acids Res* 11(13): 4287–4306.
17. Zhang XY, Hörz W (1984) Nucleosomes are positioned on mouse satellite DNA in multiple highly specific frames that are correlated with a diverged subrepeat of nine base-pairs. *J Mol Biol* 176(1):105–129.
18. Fischer TC, Groner S, Zentgraf U, Hemleben V (1994) Evidence for nucleosomal phasing and a novel protein specifically binding to cucumber satellite DNA. *Z Naturforsch C* 49(1-2):79–86.
19. Vershinin AV, Heslop-Harrison JS (1998) Comparative analysis of the nucleosomal structure of rye, wheat and their relatives. *Plant Mol Biol* 36(1):149–161.
20. Hasson D, et al. (2013) The octamer is the major form of CENP-A nucleosomes at human centromeres. *Nat Struct Mol Biol* 20(6):687–695.
21. Gent JI, et al. (2011) Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning. *Epigenetics Chromatin* 4:3.
22. Satchwell SC, Drew HR, Travers AA (1986) Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* 191(4):659–675.
23. Segal E, et al. (2006) A genomic code for nucleosome positioning. *Nature* 442(7104): 772–778.
24. Ioshikhes I, Hosid S, Pugh BF (2011) Variety of genomic DNA patterns for nucleosome positioning. *Genome Res* 21(11):1863–1871.
25. Dong F, et al. (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc Natl Acad Sci USA* 95(14):8135–8140.
26. Cheng Z, et al. (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* 14(8):1691–1704.
27. Nagaki K, et al. (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* 36(2):138–145.
28. Flores O, Orozco M (2011) nucleR: a package for non-parametric nucleosome positioning. *Bioinformatics* 27(15):2149–2150.
29. Schones DE, et al. (2008) Dynamic regulation of nucleosome positioning in the human genome. *Cell* 132(5):887–898.
30. Chung HR, et al. (2010) The effect of micrococcal nuclease digestion on nucleosome positioning data. *PLoS ONE* 5(12):e15754.
31. Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S (2011) Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci USA* 108(45): 18318–18323.
32. Krassovsky K, Henikoff JG, Henikoff S (2012) Tripartite organization of centromeric chromatin in budding yeast. *Proc Natl Acad Sci USA* 109(1):243–248.
33. Valouev A, et al. (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res* 18(7): 1051–1063.
34. Valouev A, et al. (2011) Determinants of nucleosome organization in primary human cells. *Nature* 474(7352):516–520.
35. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23(21): 2947–2948.
36. Lee HR, Neumann P, Macas J, Jiang J (2006) Transcription and evolutionary dynamics of the centromeric satellite repeat *CentO* in rice. *Mol Biol Evol* 23(12):2505–2520.
37. Prytkova TR, Zhu X, Widom J, Schatz GC (2011) Modeling DNA-bending in the nucleosome: Role of AA periodicity. *J Phys Chem B* 115(26):8638–8644.
38. Yan H, et al. (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. *Plant Cell* 17(12):3227–3238.
39. Cockell M, Rhodes D, Klug A (1983) Location of the primary sites of micrococcal nuclease cleavage on the nucleosome core. *J Mol Biol* 170(2):423–446.
40. Conde e Silva N, et al. (2007) CENP-A-containing nucleosomes: Easier disassembly versus exclusive centromeric localization. *J Mol Biol* 370(3):555–573.
41. Sekulic N, Bassett EA, Rogers DJ, Black BE (2010) The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. *Nature* 467(7313):347–351.
42. Panchenko T, et al. (2011) Replacement of histone H3 with CENP-A directs global nucleosome array condensation and loosening of nucleosome superhelical termini. *Proc Natl Acad Sci USA* 108(40):16588–16593.
43. Tanaka Y, et al. (2005) Human centromere protein B induces translational positioning of nucleosomes on alpha-satellite sequences. *J Biol Chem* 280(50):41609–41618.
44. Henikoff S, Furuyama T (2012) The unconventional structure of centromeric nucleosomes. *Chromosoma* 121(4):341–352.
45. Albert I, et al. (2007) Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* 446(7135):572–576.
46. Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* 293(5532):1098–1102.
47. Tóth K, et al. (2013) Histone- and DNA sequence-dependent stability of nucleosomes studied by single-pair FRET. *Cytometry A* 83(9):839–846.
48. Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191(4227):528–535.
49. Ventura M, Archidiacono N, Rocchi M (2001) Centromere emergence in evolution. *Genome Res* 11(4):595–599.
50. Wong LH, Choo KH (2001) Centromere on the move. *Genome Res* 11(4):513–516.
51. Tek AL, Kashihara K, Murata M, Nagaki K (2011) Functional centromeres in *Astragalus sinicus* include a compact centromere-specific histone H3 and a 20-bp tandem repeat. *Chromosome Res* 19(8):969–978.
52. Zhang H, Dawe RK (2012) Total centromere size and genome size are strongly correlated in ten grass species. *Chromosome Res* 20(4):403–412.
53. Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics* 157(3):1293–1298.
54. Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. *J Biol* 3(4):18.
55. Schueler MG, Swanson W, Thomas PJ, NISC Comparative Sequencing Program, Green ED (2010) Adaptive evolution of foundation kinetochore proteins in primates. *Mol Biol Evol* 27(7):1585–1597.
56. Hirsch CD, Wu Y, Yan H, Jiang J (2009) Lineage-specific adaptive evolution of the centromeric protein CENH3 in diploid and allotetraploid *Oryza* species. *Mol Biol Evol* 26(12):2877–2885.
57. Dawe RK, Henikoff S (2006) Centromeres put epigenetics in the driver's seat. *Trends Biochem Sci* 31(12):662–669.
58. Zhang W, et al. (2012) High-resolution mapping of open chromatin in the rice genome. *Genome Res* 22(1):151–162.
59. Sakai H, et al. (2013) Rice Annotation Project Database (RAP-DB): An integrative and interactive database for rice genomics. *Plant Cell Physiol* 54(2):e6.
60. Kawahara Y, et al. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4.
61. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
62. Marçais G, Kingsford C (2011) A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* 27(6):764–770.

Supporting Information

Zhang et al. 10.1073/pnas.1319548110

SI Text

The *CentO* consensus sequence is as follows: AACATCGCACCCACGTGTGCCAATATTGGCATTAATTGACAAAAGTTCGCCGCGCGAATCACGAAGTGAGTTTTTGCCACGAACGCACCCAATACACTCCAATATGTCCAAAAATCATGTTTTGGTGCTTTTTGAACTTTTTCATTCCGGTCAAA.

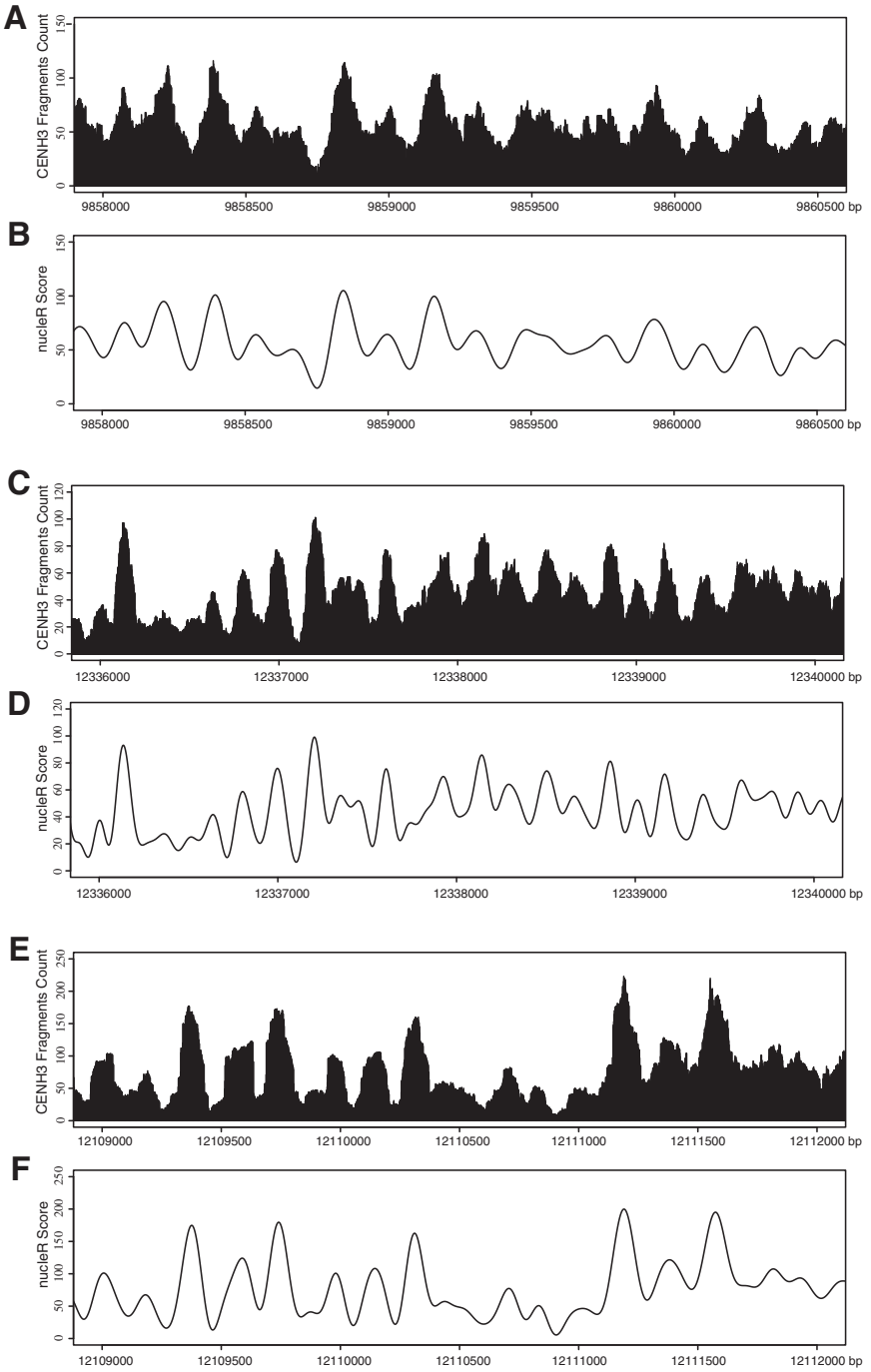

Fig. S1. ChIP-Seq fragment counts in selected regions showing positioned CENH3 nucleosomes in *Cen4*, *Cen5*, and *Cen7*. (*A*) Fragment counts in *Cen4* region (9,858,000–9,860,500 bp). (*B*) NucleR scores in the same region in *A*. (*C*) Fragment counts in *Cen5* region (12,336,000–12,340,000 bp). (*D*) NucleR scores in the same region in *C*. (*E*) Fragment counts in *Cen7* region (12,109,000–12,112,000 bp). (*F*) NucleR scores in the same region in *E*.

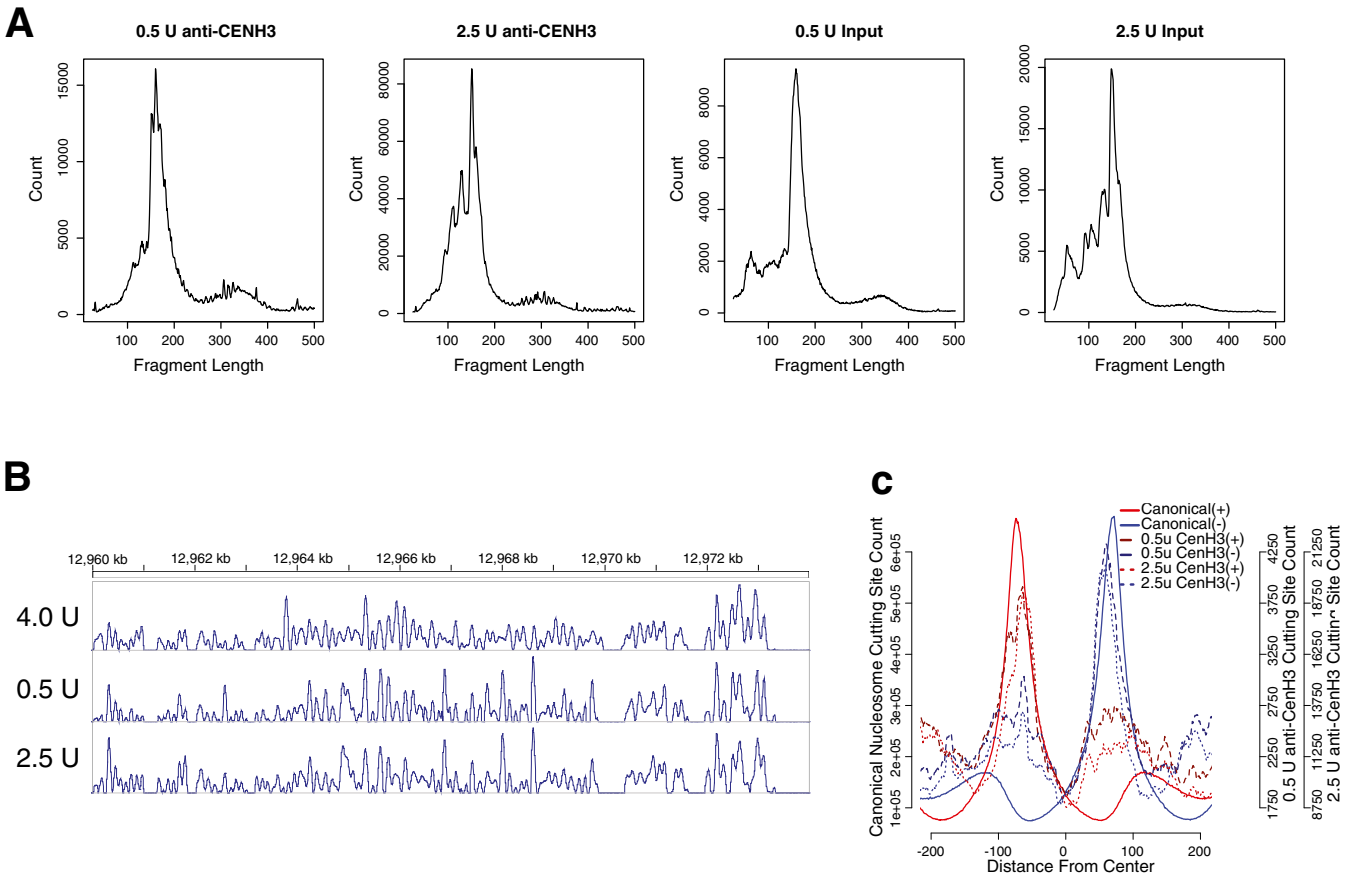

Fig. S2. Libraries made from lightly digested chromatin. (*A*) Fragment-length distribution of libraries. Fragment lengths (*x* axes) and corresponding counts (*y* axes) are shown for the 0.5 U and 2.5 U MNase-treated anti-CENH3 and input libraries. (*B*) NucleR scores within a 13-kb region of *Cen8* (12,960,000 bp–12,973,000 bp) using ChIP-seq fragments from three independent ChIP experiments with 4.0, 0.5, or 2.5 U of MNase. (*C*) Core sizes of CENH3 and canonical nucleosomes using ChIP-seq fragments from chromatin lightly digested with 0.5 U or 2.5 U MNase. The nucleosome cores are anchored by mapping MNase cutting sites from (+) and (–) strand reads (red and blue, respectively). The *x* axis represents the distance from the center. The *y* axes represent read counts from the MNase cutting sites.

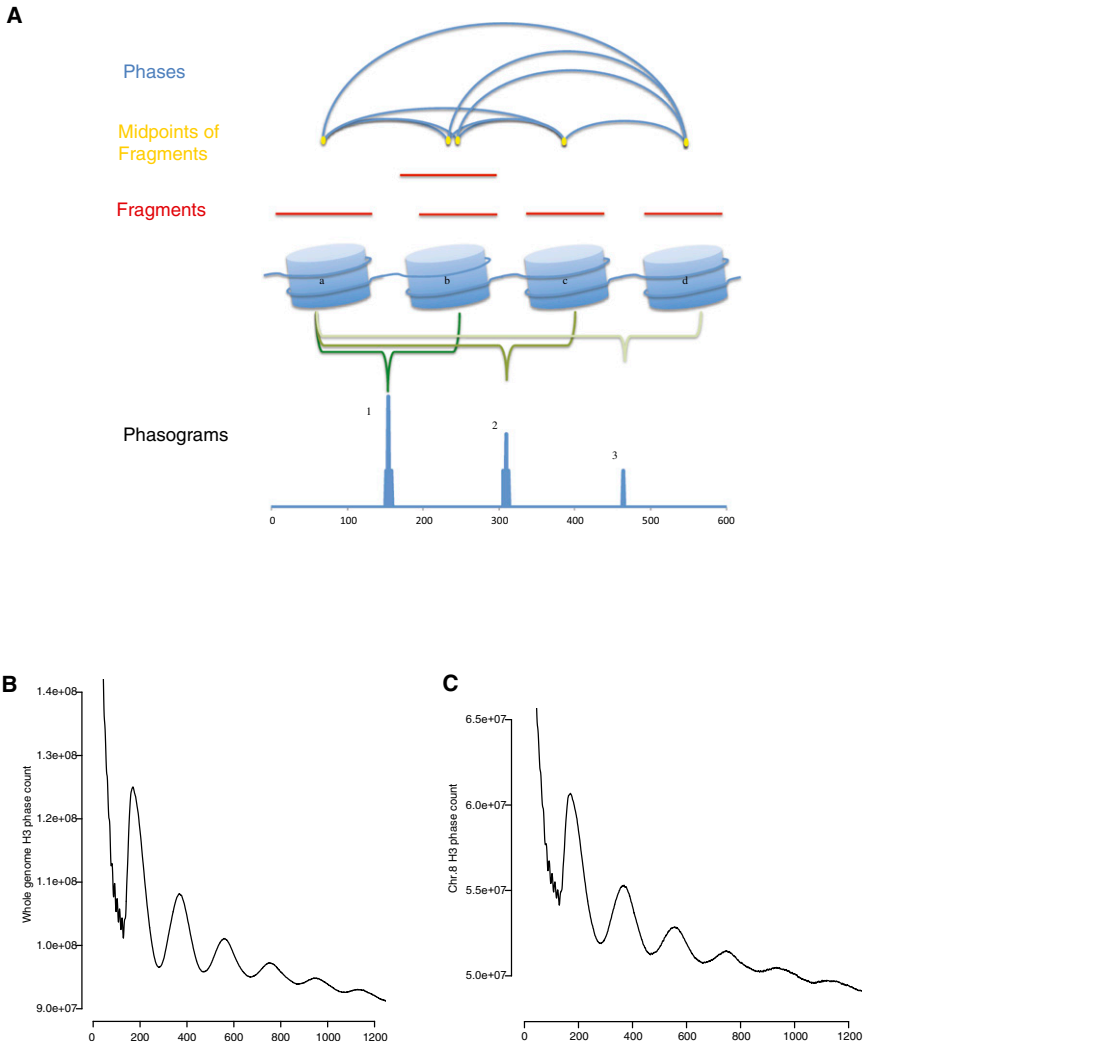

Fig. S3. Phasing of canonical nucleosomes. (*A*) Model of phasogram calculation. Phasograms are used to reveal consistently spaced nucleosomes. Blue arcs represent the distance between the midpoints of fragments. Each yellow dot indicates the midpoint of a fragment. Red lines represent fragments. The histogram represents phase frequencies. Peak no. 1 represents the distance between two-neighboring nucleosomes: a and b, b and c, c and d; peak no. 2 represents the distance between nucleosomes a and c, b and d; peak no. 3 represents the distance between nucleosomes a and d. (*B*) A phasogram generated using all input mononucleosomes, in which canonical H3 nucleosomes predominate. (*C*) A phasogram generated using all *Chromosome 8*-associated input DNA fragments excluding those from *Cen8*. The *x* axes show the range of recorded distances between the two midpoints of two fragments. The *y* axes represent the frequencies of each distance.

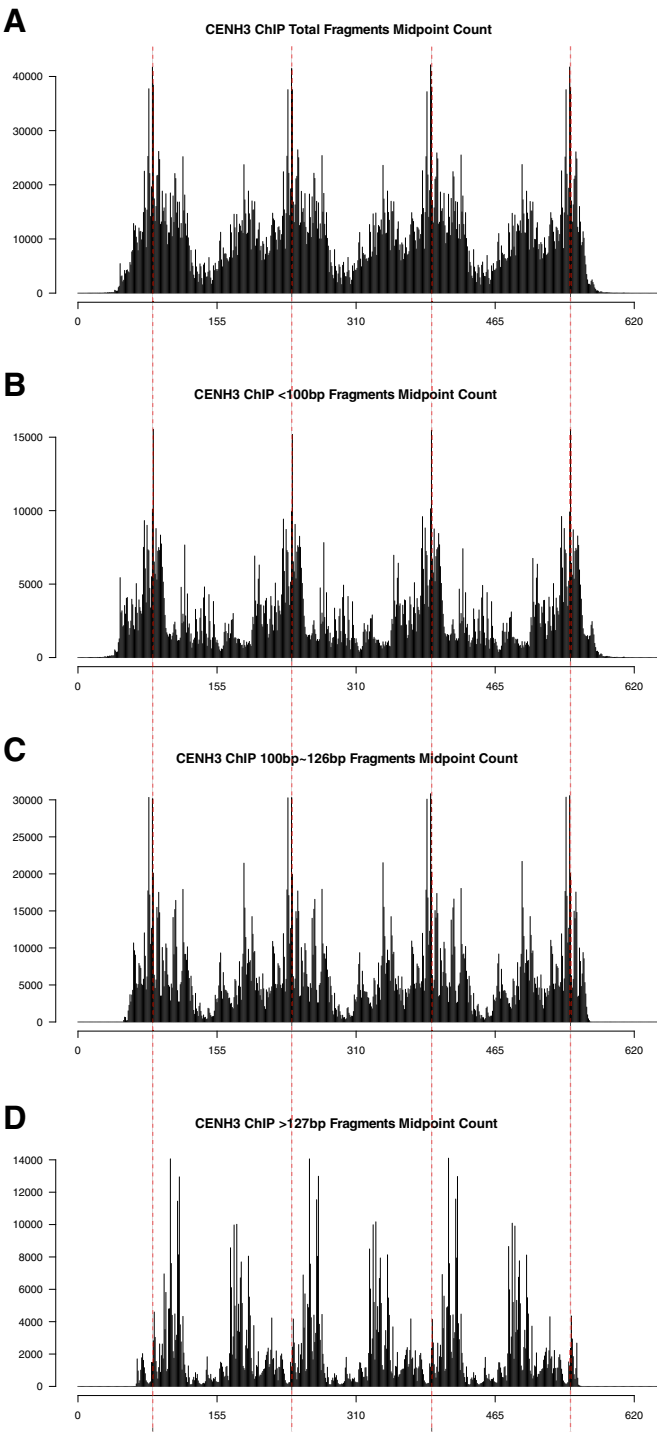

Fig. S4. Midpoint counts of *CentO* fragments. Counts (*y* axis) of midpoints of all fragments from the 4.0 U anti-CENH3 library were mapped to four copies of the *CentO* consensus sequence (*x* axis) using Novoalign. The midpoint-count peak of the total fragments is marked with a vertical dashed line on each copy of the *CentO* consensus. Shown are midpoint counts of (*A*) total fragments, (*B*) <100-bp fragments, (*C*) 100- to 126-bp fragments, and (*D*) >127-bp fragments.

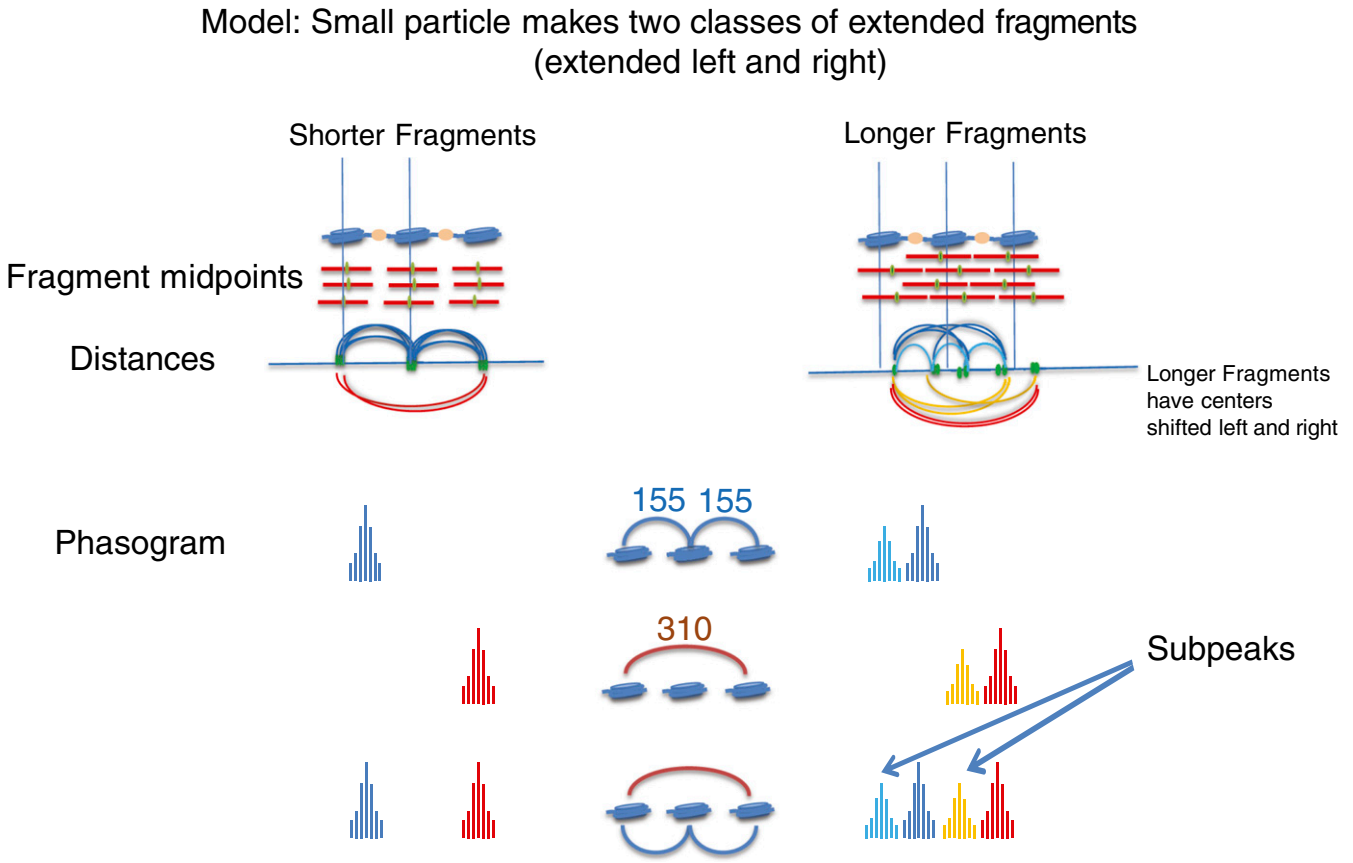

Fig. S5. Model of phasogram subpeak formation. CENH3 nucleosomes (blue cylinders) protect fragments of near-nucleosome length with midpoints near the nucleosome centers. In longer fragments, both the CENH3 nucleosome and one or more additional particles (beige balls) protect DNA. Fragments in which the particle is on the left and right sides of the nucleosome predict that the fragment midpoints are shifted left and right relative to the shorter fragments, and account for the peaks and subpeaks in the phasogram.

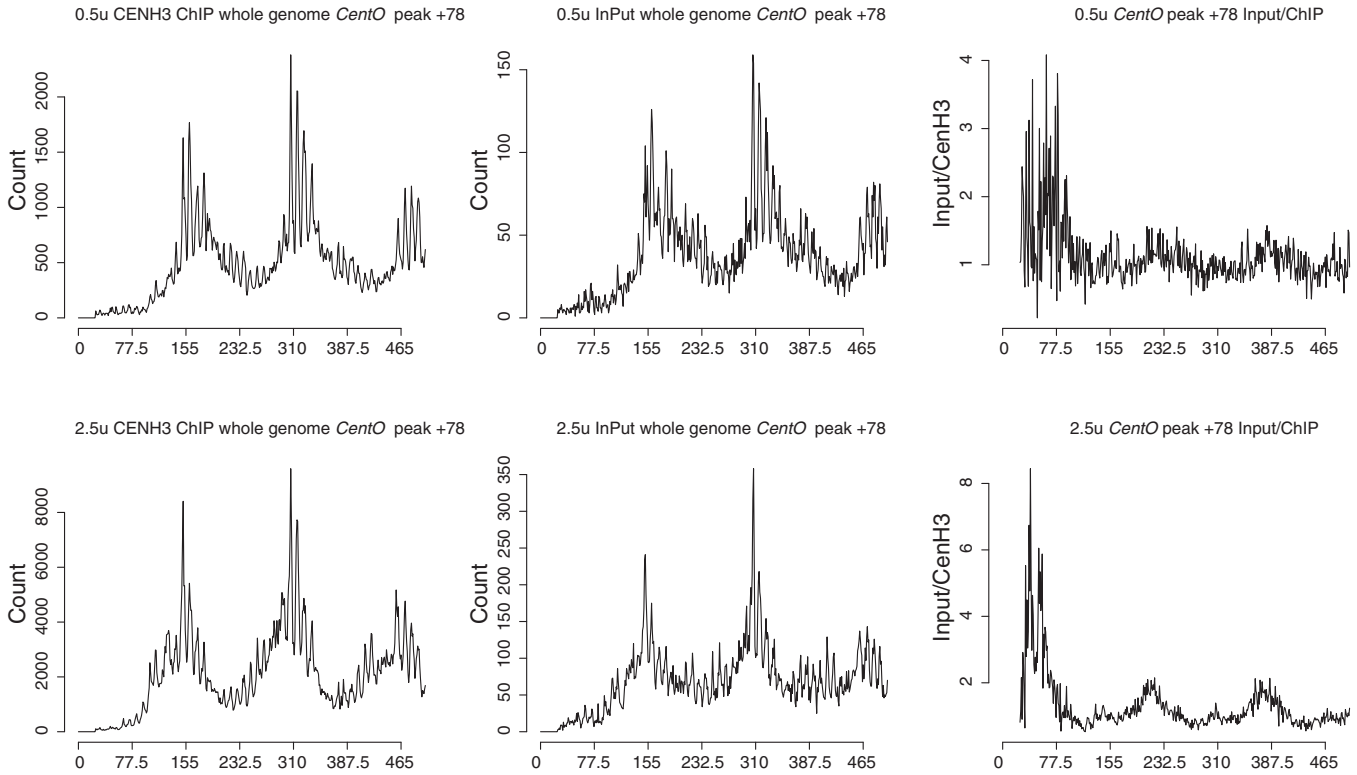

Fig. S6. Fragment-length distribution of genome-wide *CentO* fragments that pass through the halfway point between nucleosome peaks (peak + 78 bp). The proportion of <70-bp fragments (including 35-bp fragments) is increased in input libraries compared with anti-CENH3 libraries.

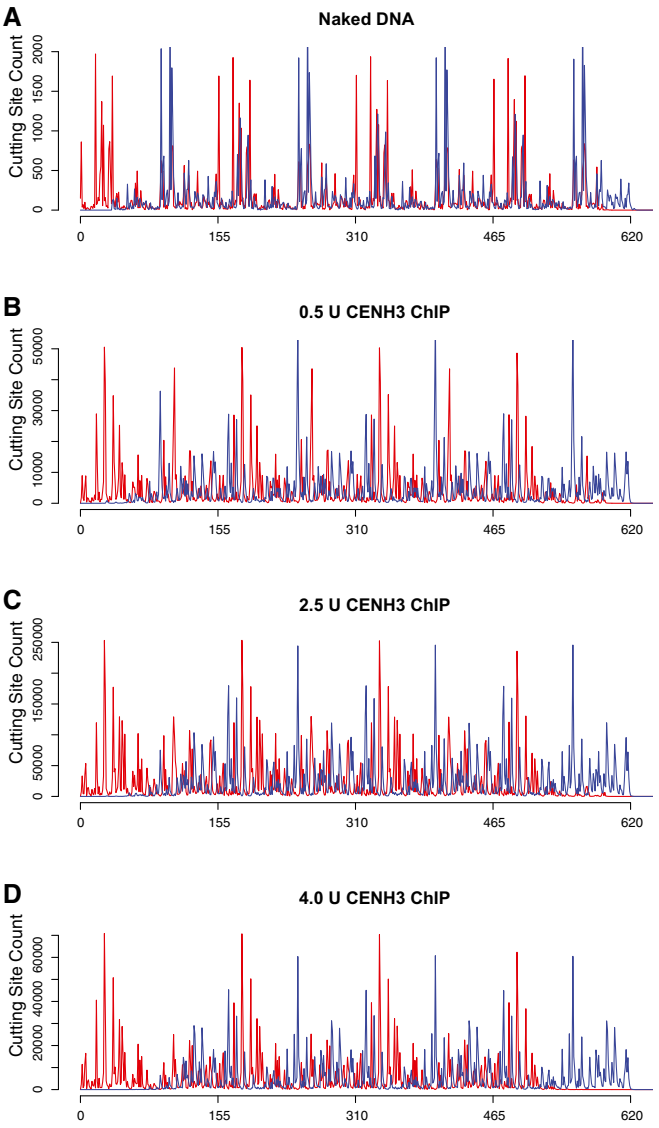

Fig. S7. MNase cutting sites in *CentO* in CENH3 chromatin and genomic DNA. Shown are MNase cutting sites on (+) and (–) strands (red and blue, respectively) of (*A*) naked genomic DNA, and of the anti-CENH3 ChIP-Seq libraries using (*B*) 0.5 U MNase, (*C*) 2.5 U MNase, and (*D*) 4.0 U MNase. The *x* axis is four copies of the *CentO* consensus sequence. The *y* axis is cutting site counts.