1550–1560 *Nucleic Acids Research, 2012, Vol. 40, No. 4* *Published online 4 November 2011*
doi:10.1093/nar/gkr862

Strong epigenetic similarity between maize centromeric and pericentromeric regions at the level of small RNAs, DNA methylation and H3 chromatin modifications

Jonathan I. Gent[1], Yuzhu Dong[2], Jiming Jiang[2] and R. Kelly Dawe[1,3,*]

[1]Department of Plant Biology, University of Georgia, Athens, Georgia, 30602, USA, [2]Department of Horticulture, University of Wisconsin–Madison, Madison, Wisconsin, 53706, USA and [3]Department of Genetics, University of Georgia, Athens, Georgia, 30602, USA

Received June 14, 2011; Revised September 22, 2011; Accepted September 26, 2011

ABSTRACT

Both kinetochore function and sister chromatid cohesion can depend upon pericentromere chromatin structure, and factors associated with heterochromatin have been proposed to have general, conserved roles in distinguishing centromeres and pericentromeres and in conferring pericentromere-intrinsic functions. We applied genome-wide sequencing approaches to quantify RNA expression, DNA methylation and histone modification distributions in maize (*Zea mays*), focusing on two maize chromosomes with nearly fully sequenced centromeres and pericentromeres. Aside from the presence of the Histone H3 variant common to all centromeres, Centromeric Histone H3 (CENH3), we found no RNA expression or chromatin modifications that clearly differentiate pericentromeres from either centromeres or from chromosome arms, nor did we identify an epigenetic signature that accurately predicts CENH3 location. RNA expression and chromatin modification frequencies were broadly associated with distance from centromeres, gradually peaking or dipping toward arms. When interpreted in the context of experimental data from other systems, our results suggest that centromeres may confer essential functions (such as cohesion retention) to flanking sequence regardless of the local heterochromatin profile.

INTRODUCTION

The sequence of steps in chromosome movement during cell division depends on a unique chromatin environment in centromeres. Many factors can contribute to the critical attributes of centromere chromatin, but the histone variant CENH3 (also referred to as CENP-A) is viewed as a defining feature of an active centromere [reviewed in (1,2)]. In organisms where centromeres exist within a single domain on each chromosome, the adjacent domains are generally referred to as pericentromeres. Pericentromeres also have essential roles in chromosome movement during cell division as they provide the framework for keeping sister chromatids oriented correctly until anaphase. Between prophase and anaphase, sister chromatids remain attached along their pericentromeres specifically, as cohesin is removed everywhere else [reviewed in (3)]. Various factors associated with heterochromatin during interphase have been proposed to set the stage for the retention of cohesin at pericentromeres, including Heterochromatin Protein 1 (HP1), trimethylation of lysine 9 on histone H3 (H3K9me3) and RNAi (4,5). Transient deposition of histone modifications during early mitosis has also been proposed to function in this regard, such as phosphorylation of serines 10 and 28 on histone H3 and phosphorylation of serine 121 on histone H2A (6,7).

Extensive investigation in the fission yeast *Schizosaccharomyces pombe* pericentromere chromatin has led to the discovery that specific chromatin factors and an RNAi-related mechanism are key not only for proper cohesin dynamics but also for proper placement of CENH3 (8,9). The DNA composition itself contributes to pericentromere function, as repetitive sequences in the pericentromere are transcribed and give rise to siRNAs that recruit heterochromatin factors (5,10,11). Despite the comparatively simple composition and small size of *S. pombe* centromeres and pericentromeres, similarities have been identified in the much more complex centromeric regions of diverse plants and animals. In addition to heterochromatin factors in pericentromeres, histone modifications associated with euchromatin have been observed in centromeres of both *S. pombe* and animals (12,13).

*To whom correspondence should be addressed. Tel: +706 542 1658; Fax: +706 542 1805; Email: kelly@plantbio.uga.edu

Schizosaccharomyces pombe is different from many other eukaryotes in that it lacks DNA methylation, which is an important component of heterochromatin in organisms which are capable of it. Hence extrapolation from *S. pombe* suggests that DNA methylation could contribute to pericentromere heterochromatin function (though not actually in *S. pombe* itself).

The similarities between *S. pombe* and other organisms have led to much conjecture about a conserved, functional connection between pericentromere heterochromatin and both centromere function and sister chromatin cohesin (14,15). However, the extent to which the *S. pombe* model applies to other organisms, with its reliance on control elements intrinsic to pericentromeres, is unclear [see discussions in (6,16,17) and references therein]. These arguments against broad applicability include: (i) a general lack of mitotic centromere or cohesin defects in RNAi and chromatin modification mutants in plants and animals; (ii) Profound differences in both DNA structure and siRNA expression patterns between *S. pombe* and other organisms; (iii) Dispersed rather than pericentromere-concentrated heterochromatin marks in plants with large genomes; (iv) Existence of functional centromeres that lack canonical pericentromere sequences or even heterochromatin domains (18,19); and (v) Lack of a clear link between HP1 and centromere function in animals (20–22) and an established euchromatic function for the HP1 homolog in plants (23–25). These observations support the notion that the predominant controlling factors are derived from within the centromere itself, both for internal control (CENH3-containing chromatin) and external (pericentromere cohesion) (17).

Pericentromere-intrinsic control models derived from *S. pombe* predict that the pericentromere chromatin should be functionally and visibly distinct both from the centromere chromatin and the arm chromatin. The presence of CENH3 within the centromere specifically is one such distinction; however, centromeres also contain measurable amounts of cannonical H3 and of course H2A, H2B and H4 [reviewed in (1,2)]. Within the scope of histone modifications, DNA methylation, and RNA expression patterns, one would expect to see a pericentromere chromatin signature that could be connected to its functional role. Such a signature is clearly evident in *S. pombe* and others (12). Maize presents a useful organism for direct testing of this expectation because of its large, sequenced centromeres and high-resolution maps of CENH3 occupancy (26,27) and because of its rich history of genetic and cytological analysis of chromosome structure. Previous cytological analyses in maize have revealed that its heterochromatin is dispersed along the length of its chromosomes rather than being highly concentrated just at pericentromeres, and the euchromatin histone modification H3K4me2 that is enriched in *S. pombe* and animal centromeres shows no enrichment in maize centromeres (28–30).

In order to more fully characterize RNA expression and chromatin patterns in maize centromere regions and to search for evidence of a pericentromere chromatin signature, we examined siRNA and poly(A) RNA expression, DNA methylation, and four euchromatic histone modifications patterns—H3K4me3, H3K36me3, H3K27me3 and acetylation of H3K9—across chromosomes in maize using both our own and publicly available deep sequence data. We found no evidence for anything peculiar to pericentromeres, but instead found that the centromere itself was the predominant feature in terms of large-scale patterns. The general trend manifest from these analyses was the centromere rather than pericentromere providing the most visible mark along the chromosome, representing either the top or bottom of a large curve with opposite extremes toward the arms.

MATERIALS AND METHODS

Derivation of small RNA sequencing libraries

Wild-type maize seeds, B73 stock, were germinated in vermiculite at 30°C for 3 days and the last 2–3 mm of the primary root tips cut off with a razor blade and frozen on dry ice. After grinding to a fine powder in liquid nitrogen, small RNA was extracted using the mirVana™ miRNA Isolation Kit (Ambion). A Solexa/Illumina library was then prepared using oligos as described in (31) and a ligation scheme derived from (32) and (33), which selectively captures Dicer products and other small RNAs with both a 3′-OH and a 5′-monophosphate. In order to distinguish this sample from others also included in the same sequencing run, we included a barcode CGT on the 5′-adapter. Samples were sequenced with the Illumina Genome Analyzer II. The 5′-barcode and 3′-adapter sequence were trimmed off as described previously except that the 3′-adapter was identified and trimmed based on a perfect match to the first 8 nt of the adapter (CTGTAGGC) (31). The raw, untrimmed reads are available at the NCBI Sequence Read Archive, accession SRR218319.1.

Derivation of CENH3-chromatin immunoprecipitation bisulfite sequencing libraries

Chromatin immunoprecipitation (ChIP) was carried out as previously described (34) using anti-CENH3 antibodies (35). The 14-day-old B73 seedlings were used as the source tissue. The ChIPed DNA was treated with sodium bisulfite and prepared for Illumina sequencing using a method similar to that of Lister *et al.* (36). Reads are available at the NCBI Sequence Read Archive, accession SRR218318.1.

Read processing and alignments

After removal of barcodes and adapters from the small RNA reads, known microRNAs were filtered out by Blasting against a set of maize mature miRNA sequences obtained from miRBase (release 16) [www.mirbase.org; (37)]. We used blastall with default parameters except the expectation value (*E*) was set to 1 e–6. The resulting miRNA-filtered reads, which we refer to as siRNA reads, were sorted by read length to produce individual fasta files corresponding to each length from 20 to 26 nt. Each file was then aligned to the complete maize genome version 2, obtained from www.maizesequence.org. Alignments were

carried out with bowtie software (38). For analysis of uniquely-aligning reads, i.e. those with a single best alignment, the following parameters were used: '-n 0 -l 20 -m 1 –best'. For analysis of total aligning reads, the parameters were: '-n 0 -l 20 -M 1 –best'. Alignment of the 454-sequenced CENH3 reads (27) to the reference genome to mark the centromeres in figures was carried out with bowtie similarly to siRNA reads except that only uniquely-aligning reads were analyzed and a longer seed length was used '-n 0 -l 100 -m 1 –best'.

Control reads were derived from a sample of randomly-fragmented B73 genomic DNA sequenced with Illlumina paired-end technology (39). We combined the reads from both ends into a single set, which we treated as single-end, trimmed to the appropriate length (20, 26, or 36 nt), and processed identically to the siRNA reads (except without removal of miRNA matches). This control was intended primarily as an indicator for domain-level effects from errors in the physical map, particularly omissions, as a large proportion of centromere repeats have not yet been placed on the physical map. However, any biases inherent to Illumina sequencing may also be reflected in this sample.

Bisulfite reads were aligned to the genome and to individual repeat sequences using BS Seeker (40). Only reads that could be unambiguously assigned to a single locus with no mismatches were used for quantifying methylation status, using default parameters except '-t N -e 36 -m 0'. Control, non-bisulfite-treated reads from Tenaillon *et al*. (39) were processed exactly the same. The low error reads were derived by extracting only the first 36 nt; the high error reads were from nucleotides 49 to 84 of the reads.

Poly(A) enriched, ChIP reads (35 nt) and the control reads (low error reads; see above) were processed similarly to the siRNA reads, but without removal of miRNA matches.

RESULTS

siRNA expression patterns

Given the importance of siRNAs in directing histone modifications in *S. pombe* pericentromeres and in directing DNA methylation of repetitive elements in plants, we first surveyed siRNA expression patterns genome-wide for evidence of a pericentromeric chromatin signature. We prepared Illumina small RNA sequencing libraries from RNA extracted from maize root tips (which contain a large fraction of dividing cells) and aligned the resulting siRNA reads onto the maize genome. A complication with aligning short reads to a highly repetitive genome is that in many cases, it is impossible to determine which of the multiple identical loci is the source of a read. Two common solutions are to disregard all but unique reads (those that can unambiguously be assigned to a single locus), or to include all reads and to assign a locus at random from the possible options. One method loses information, especially when dealing with highly repetitive centromeres; the other entails a higher level of misinformation, since it will assign a portion of repetitive reads to the wrong loci. We opted to use both methods, as displayed in Figure 1 and Supplementary Figure S1. Since one of the distinguishing characteristics of functionally distinct types of siRNAs is length, we split the siRNA reads into sets based on length, from 20 to 26 nt. The most abundant, and potentially most interesting because of their connection to heterochromatin, are the 22-nt and 24-nt types (41,42)

Centromeres were defined to include all the chromatin within the outer bounds of the region of the chromosomes enriched for CENH3 [defined by ChIP with CENH3 antibodies (27)]. In all but two chromosomes (chromosomes five and seven) a single, MB-scale domain of CENH3 was designated as the centromere in a prior study (27). These regions are also referred to as centromere cores, designating that they are defined precisely by CENH3 occupancy. Although centromeric sequences are highly repetitive, there is still enough variation in the DNA content to align even short reads uniquely, especially in centromeres two and five, which have been traversed and sequenced (27). For the 22-nt siRNAs, 11% of the putative centromere matching reads were unique. Small RNAs have also been successfully mapped to rice centromere 3 (43). We note that in some rice centromere cores, CENH3 domains are interspersed with extended regions of H3-containing chromatin (44). This is probably true in maize as well, but at this stage we lack the resolution to determine where such regions may lie.

The 22-nt siRNAs were observed at a consistently low baseline level across chromosome domains, with large spikes in specific regions (these are not known microRNAs). In contrast, the chromosomal distribution of 24-nt siRNAs revealed a pattern of higher density toward the arms and a low point within the centromere, reminiscent of both recombination frequency and gene density (26,45,46). On chromosome 5 there was a particularly low density of 24-nt siRNAs over the region associated with CENH3, however a similar dip was not observed on chromosome 2. We also examined the chromosomal distribution of other lower abundance siRNAs between 20 and 26 nt on chromosome 2; however, none showed anything unusual in the pericentromeric region (Supplementary Figure S1A and B). These high-resolution analyses of siRNAs within centromere core regions and surrounding areas provide additional evidence that centromeric domains are compatible with both transcription and siRNA production, but only at a low level in both centromere and pericentromere relative to chromosome arms in root tip cells.

DNA methylation patterns

The lack of a pericentromeric 24-nt siRNA signature suggested that associated DNA methylation might also lack a distinguishing pericentromeric pattern. DNA methylation in plants exists in three forms: cytosine methylation in CG, CHG and CHH contexts, where H is anything but G. The CHH type is most heavily dependent upon 24-nt siRNAs [reviewed in (47)]. Our intent was to focus on the centromere core regions, which had been shown in some cases to display hypomethylation in prior

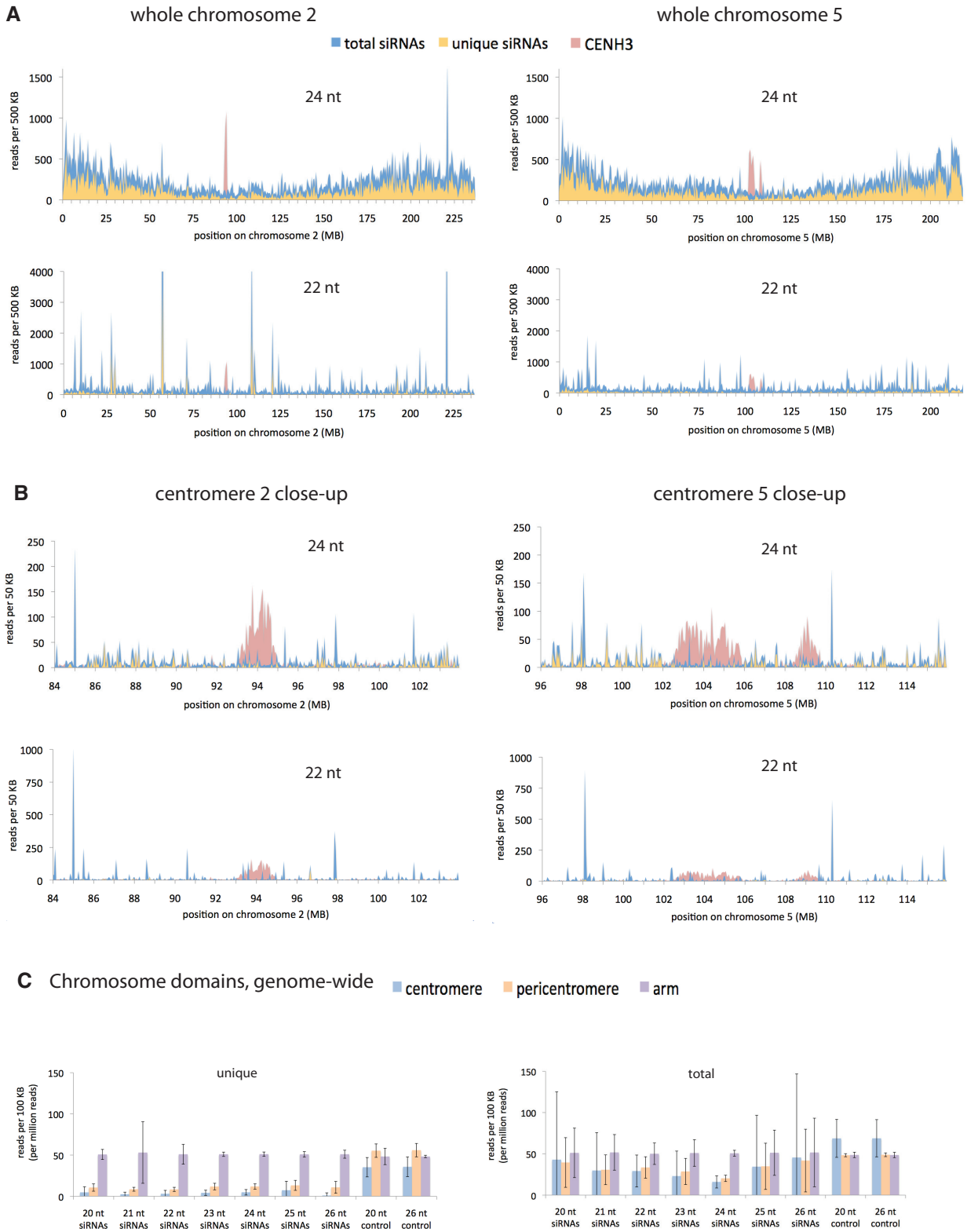

Figure 1. Comparison of siRNA abundance across chromosome domains. (**A** and **B**) Chromosome-wide views and centromere close-ups of 24- and 22-nt siRNAs. (A) Whole chromosome views of chromosomes 2 and 5, and (B) close up views of the centromeres plus 5 Mb of sequence on either side. For small RNAs, the Y-axes show read counts, normalized per million total genomic reads for each length at each 500-kb or 50-kb interval. Unique siRNAs are those corresponding to a single locus whereas total siRNAs include those which could be derived from multiple loci (one of which was chosen at random). Note that some peaks extend off the chart. The CENH3 reads mark unique loci within centromeres and are derived from CENH3 ChIP [described in (27)]. CENH3 read count was normalized per hundred thousand unique genomic reads. See also Supplementary Figure S1. (**C**) Average siRNA count for each chromosome domain. Centromeres are the domains occupied by CENH3, as defined by the CENH3 ChIP reads. For the purposes of this study, pericentromeres extend 5 MB on either side of the centromeres, and arms are the rest of the chromosomes beyond the pericentromeres. Error bars are standard deviations arising from differences in the values from the 10 chromosomes. The control reads are derived from an Illumina-sequenced, randomly-fragmented genomic DNA sample (39) and are included to provide an estimate of the bias introduced in each domain due to inaccuracies in the physical map and biases arising from Illumina sequencing.

immunostaining studies using methyl-cytosine-specific antibodies (48–51). To this end, we first performed ChIP using maize CENH3 antibodies to enrich for centromeric sequences. We then treated the sample with bisulfite and sequenced it with Illumina technology. The experiment conducted in this way clearly delineated the boundaries of the centromere cores while retaining sufficient read depth to cover the rest of the maize genome (Figure 2 and Supplementary Figure S2A and B). The analysis revealed gentle, chromosome-level trends of CG and CHG methylation increasing toward centromeres and decreasing toward arms, as has been observed in many other species (52,53). In contrast, CHH methylation was generally similar across different domains of the chromosomes as has been observed previously in rice (52). Although overall methylation was similar in centromeres and pericentromeres, within the centromere core we noted a subtle increase in CG methylation and decrease in CHG methylation, with notable variation among centromeres (Figures 2, Supplementary Figure S2A and B). Such variation may be attributable to varying sequence coverage and/or the relative sizes of the *CentC* arrays in different centromeres, which range from <300 kb to several megabases (27,54).

This genome-wide analysis was initially limited to unique loci in order to ensure high accuracy in quantification of methylation frequency (Supplementary Figure S2C). To assess the level of methylation within repetitive areas, which are under-presented in the genome assembly, we analyzed reads that match the tandem repeat (satellite) *CentC* and the retrotransposon *CRM2* independently of map position and copy number (Figure 2D). These two elements are the most abundant of the repetitive elements in maize centromeres (26,27). We found that CHH methylation was present at very low levels for *CentC* and *CRM2*, similar to CHH methylation at unique centromere loci (~1% of cytosines in a CHH context were methylated in each case) The average CHG methylation values for both *CentC* and *CRM2* as a whole were lower than the CHG values in unique centromeric loci (42% for *CentC* and 45% for CRM2 as compared with 65% for unique centromeric loci). These results may partially explain the prior observation that active maize centromeres stain poorly with methyl-cytosine-specific antibodies in maize and other plants (44,46). In contrast, CG methylation in centromere repeats was not notably different from either single copy pericentromeric or single copy centromeric regions. The data reveal that 88% of the CG dinucleotides within the *CentC* repeats are methylated on average (Figure 2D); however it is likely that the levels of CG methylation vary within long arrays (44–47).

mRNA and euchromatin histone modification patterns

Transcription has been observed at both centromeres and pericentromeres in many other organisms, as evidenced by sequence identity with known centromere- or pericentromere-specific repeats. In *S. pombe*, transcription driven by RNA polymerase II occurs at pericentromeres but not centromeres (11,12). Similarly, in rice, genes exist in centromere core regions but they tend to lie within subdomains of normal H3 chromatin and not the subdomains identified by CENH3 (44). Nevertheless, a direct comparison is difficult because repetitive sequences in plants can be transcribed via a complicated interaction of three RNA polymerases, II, IV and V, of which only II exists in *S. pombe* [reviewed in (55)].

Maize has the distinction of being the first organism for which centromere RNA was found to associate directly with CENH3 chromatin (56) but the most detailed analyses have been done in mammals, where transcription has been associated with both chromatin modifications and kinetochore integrity [reviewed in (57)]. In order to quantify RNA expression over the entire centromere and pericentromere in maize, we examined a previously published set of poly(A)-enriched RNA fragments (58). We took the raw, unfiltered data from this study, and produced genome-wide alignments to measure steady-state RNA levels (Figure 3). The clear image that emerges from this is a general depletion of poly(A) RNA toward the centromere and enrichment in chromosome arms, as expected based on density of protein-coding genes. However, low but detectable expression was also present in centromeres.

To measure transcript abundance for the tandem repeat *CentC* and the retrotransposon *CRM2*, we counted the reads that matched to either element. Of the 26 071 222 genome-matching, poly(A)-enriched reads, 4 matched *CentC* and 318 matched *CRM2*, and only a handful of these mapped uniquely. An analysis of the non-unique *CRM2* alignments on centromere 2 suggested that most of the transcripts were derived from the centromere core areas (Supplementary Figure S3), as expected based on the distribution of *CRM2* elements (27). We also observed a substantial number of siRNAs derived from *CentC* and *CRM2*: In a sample of 1 550 399 total genome-matching siRNA reads, 127 matched *CentC* and 545 matched *CRM2* (Supplementary Figure S1C and D). These siRNAs are likely derived from non-polyadenylated or otherwise unstable templates. *CentC* transcripts of unusual sizes have also been detected by RNA blotting and in sequence libraries from non-poly(A)-enriched RNA (56,59,60). Based on the centromeric location of *CentC* DNA (27), and the physical association of *CentC* transcripts with CENH3 chromatin (56), these data provide compelling evidence for transcription from centromeres in maize, though not in a form that is readily captured and sequenced by poly(A) enrichment. In contrast, evidence for *S. pombe*-type transcription from pericentromeres is lacking in maize.

Euchromatin histone modifications are generally thought to be depleted from pericentromeres, and in both *S. pombe* and multiple animal species, certain euchromatin modifications (H3K4me2 and H3K36me3) have been reported to be enriched with the centromere itself relative to pericentromeres (12,13,61,62). Such an enrichment is not observed by immunostaining in maize (28,29), nor in the filamentous fungus *Neurospora crassa,* where H3K4me2 is absent in centromeres but enriched in flanking areas (19). The Wang *et al.*, study from which we derived our analysis of poly(A) RNA fragments also included a large-scale ChiP with antibodies for four

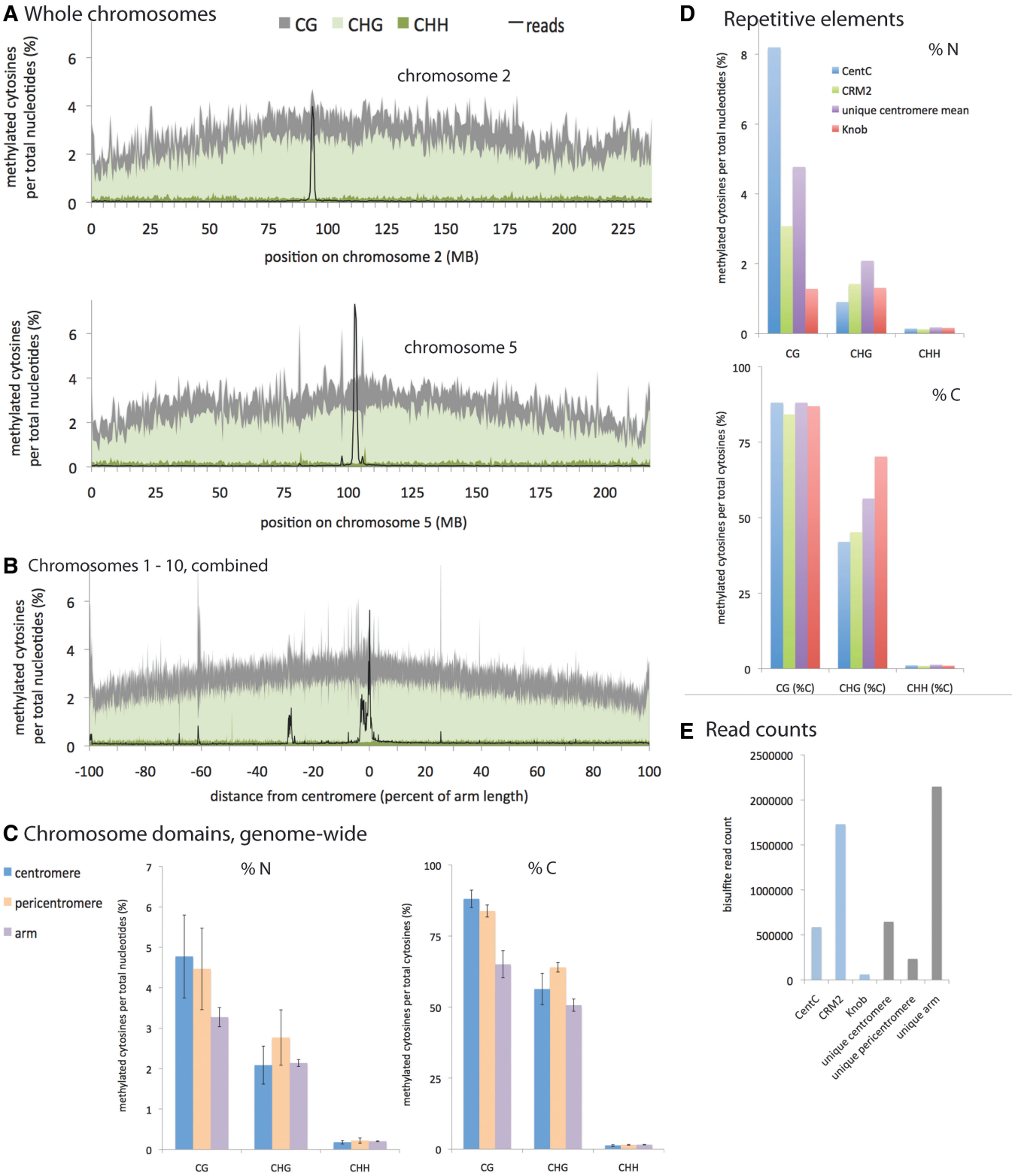

Figure 2. Comparison of DNA methylation across chromosome domains. (**A**) Chromosome-wide view of DNA methylation in each sequence context for unique loci on chromosomes 2 and 5. Each chromosome was divided into 500-kb intervals, and the methylation percentage was calculated for each interval. Also displayed (as a black line) is the number of unique bisulfite reads at each position, making the enrichment for centromeres in this ChIP sample clearly evident. To make the scale appropriate for the figure, the total read count was divided by 10 000. See also Supplementary Figure S2. (**B**) Chromosome-wide view of DNA methylation in each sequence context for unique loci for all 10 chromosomes combined. All the chromosomes were aligned at the center of their centromeres and their arm lengths normalized such that positions were converted to a percentage of total distance from the centromere (0% to 100% of arm length, *X*-axis). The arms were divided into 1000 intervals, the length of which depended on the particular chromosome. The methylation percent (Y-axis) is a cumulative value calculated from all reads on all chromosomes within a particular interval. Note that several maize centromeres contain mainly the *CentC* repeat and have been poorly sequenced, thus, the combined view (also for C

(continued)

euchromatin histone modifications. We used the reads from this study to ask whether an enrichment could be seen in centromeres, or whether any clear difference between centromeres and pericentromeres could be detected (Figure 3 and Supplementary Figure S3). While some reduction in H3 modifications is expected in centromeres by virtue of CENH3 occupancy, all four of the modifications examined—H3K4me3, H3K36me3, H3K27me3 and H3K9ac—were generally depleted within both centromeres and pericentromeres, similar to 24-nt siRNAs, poly(A) RNA, density of protein coding genes, and crossover suppression, and opposite to that of CG and CHG methylation. The chromosome landscape visible from these views is consistently the centromere at the summit of the hill or bottom of the valley, and the pericentromere occupying the slopes on either side. The landscapes are notably gentle, though highly variable, and if not for the presence of CENH3, we would find it difficult to pinpoint the locations of centromeres by standard epigenetic profiling.

DISCUSSION

Using genome-wide epigenetic profiling, we found that maize pericentromeres differed little from centromeres (with the key exception that CENH3 replaces canonical of H3 in the kinetochore domain). There are many caveats to this study relating to the repetitive nature of centromeres and the fact that only two maize centromeres are fully assembled, and none are fully sequenced base for base. The results apply most firmly to centromeres two and five where nearly complete sequences are available. In the most conservative sense, our data can be viewed as pertaining to regions that are unique by a 36-nt or less sequence tag. However, despite the short size of the reads, a large proportion can be aligned to regions within the centromere core and flanking pericentromeric areas (e.g., Figure 2E and 1C).

If arms are viewed as one extreme and centromeres as the other, pericentromeres represent an intermediate state but are much more similar to centromeres than to arms. This general trend was true for siRNAs, DNA methylation, poly(A)-enriched RNA, and for four histone modifications—H3K4me3, H3K36me3, H3K27me3 and H3K9ac. These results are similar to what has been described in *Caenorhabditis elegans*, *Neurospora crassa*, *Candida albicans*, *Saccharomyces cerevisiae* and many other species in that they lack defined pericentromeric heterochromatin regions. It is of course possible that other marks or specific combinations of marks delineate the pericentromere in these species, but such an unidentified code would not generally be classified as heterochromatin as we understand it.

Our data showing that maize heterochromatin gradually decreased in relation to the distance from the centromere differs substantially from what has been shown in several widely studied species that have an enrichment of heterochromatin directly flanking centromeres, including *S. pombe*, *Drosophila melanogaster*, *A. thaliana*, and humans. However, a close look at these systems suggest that they are dissimilar in fundamental ways [reviewed in (16,17)]. For example, HP1 performs a critical function of ensuring cohesion at *S. pombe* pericentromeres but appears to be dispensable for cohesion in humans (20–22) and in *A. thaliana* (23–25). Furthermore, centromeres are orders of magnitude smaller in *S. pombe* than in these other species. These data suggest that while heterochromatin may frequently be found in close proximity to centromeres, its role in chromosome segregation varies among species. *S. pombe* may provide an extreme example with its dependence on heterochromatin for de novo deposition of CENH3 and for proper cohesion dynamics (4,5,8,9)

In contrast to pericentromeric heterochromatin, CENH3 is always found at active centromeres, making it a prime candidate as regulator of its own deposition, although perhaps indirectly [reviewed in (1,2)]. The most compelling arguments in favor of the view that centromeres and pericentromeres can function independently comes from cases where centromeres are known to have moved to new locations. In the yeast *S. cerevisiae,* centromeres that have been empirically moved to chromosome arms are sufficient for apparently normal cohesion behavior flanking the new centromere (63). Furthermore, it is well established that human neocentromeres can form on chromosome arms and in the absence of both canonical pericentromeric DNA sequence and of visible flanking heterochromatin. Conversely, when a centromere has been epigenetically deactivated on a dicentric maize chromosome, pericentromeric phosphorylation of H3 serine 10 and cohesion are lost, demonstrating a reliance on the centromere (64). We emphasize, however, that CENH3 must interact with H3 at a local level since they are in close proximity and H3 is present in centromere cores. For example, multiple experiments with human artificial chromosomes have revealed interactions

Figure 2. Continued
below) could be biased toward the better sequenced centromeres. Several chromosomes have errors in their physical maps leading to substantial CENH3 signals on arms (black line). (**C**) Average DNA methylation in unique loci in each chromosome domain. Included are both percent of methylated cytosines per total nucleotides, as in (A), and percent of methylated cytosines per total cytosines in each context. Centromeres are the domains occupied by CENH3, as defined by the CENH3 ChIP reads. For the purposes of this study, pericentromeres extend five MB on either side of the centromeres, and arms are the rest of the chromosomes beyond the pericentromeres. Error bars are standard deviations arising from differences in the values from the 10 chromosomes. (**D**) DNA methylation in repetitive loci. *CentC* is the tandem repeat of maize centromeres, *CRM2* is the dominant centromere retrotransposon, and *Knob* is a tandem repeat found on arms. Shown are both percent of methylated cytosines per total nucleotides, as in (A), and percent of methylated cytosines per total cytosines in each context. (**E**) Abundance of bisulfite reads. Raw counts are depicted for each chromosome domain (uniquely aligning reads in gray) and for repeats (blue). Due to enrichment for CENH3 domains by ChIP prior to bisulfite treatment and sequencing, a substantial number of uniquely aligning reads come from centromeres, despite the high density of repeats in centromere repeats.

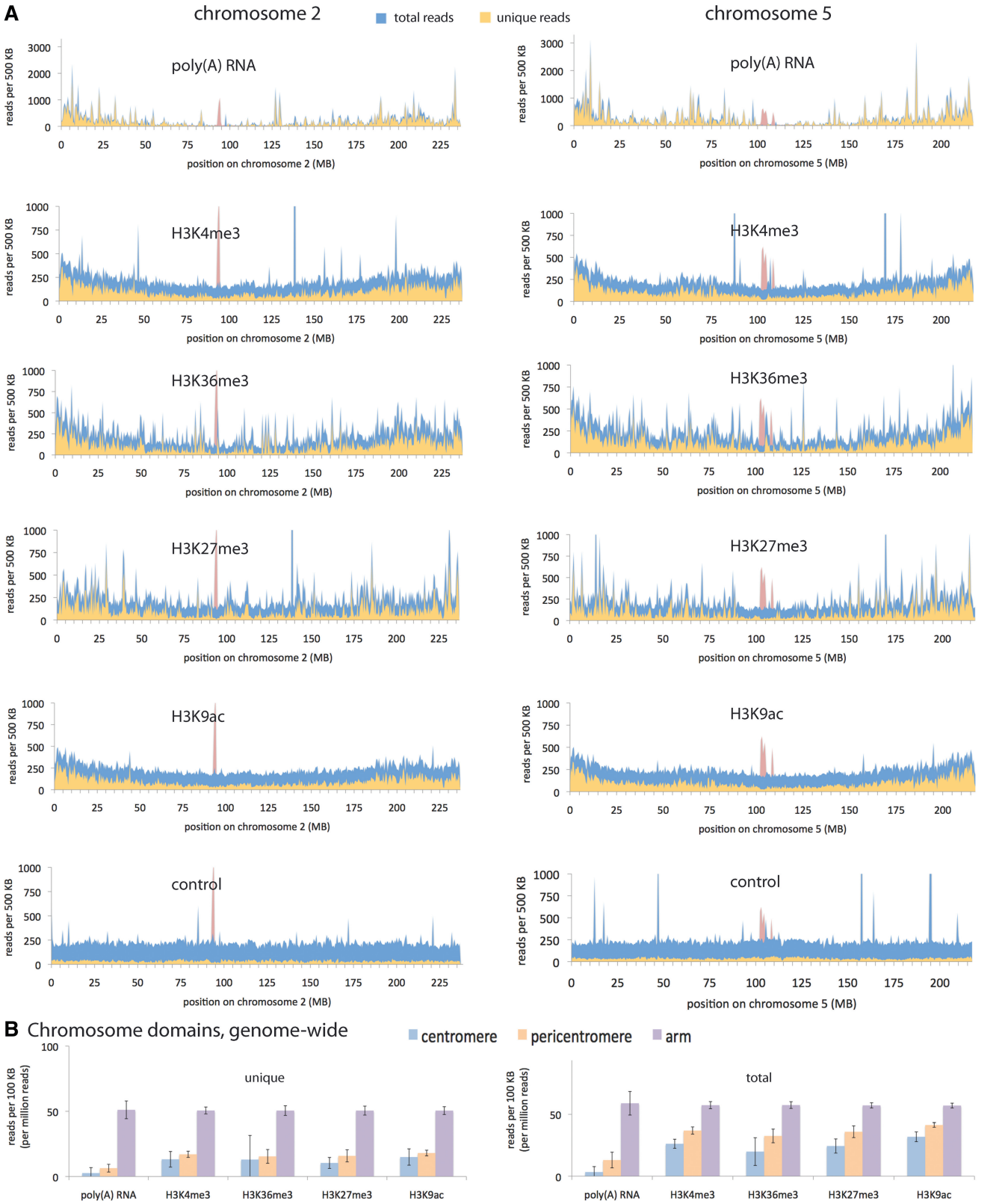

Figure 3. Poly(A) RNA and euchromatin histone modification patterns. (**A**) Chromosome-wide view of reads from poly(A) RNA fragments and from ChIP with antibodies against H3K4me3, H3K36me3, H3K27me3 and H3K9ac [derived from (58)]. Also included are a set of control, genomic DNA fragments (39). All counts are normalized per million reads. Note that some peaks extend off the chart. The CENH3 reads mark unique loci within centromeres and are derived from CENH3 ChIP [described in (27)]. See also Supplementary Figure S3. (**B**) Average read count for each chromosome domain. Centromeres are the domains occupied by CENH3, as defined by the CENH3 ChIP reads. For the purposes of this study, pericentromeres extend 5 MB on either side of the centromeres, and arms are the rest of the chromosomes beyond the pericentromeres. In order to lessen biases from errors in the physical map and Illumina sequencing, all values were normalized by the same genomic DNA fragment control reads as displayed in Figure 1C. In this case, the control reads were trimmed to 36 nt to match the read length of the poly(A)-enriched RNA and histone modification reads. For the unique alignments, this normalization also removed the bias against repetitive regions, since the control reads were subject to the same unique alignment criteria. Error bars are standard deviations arising from differences in the values from the 10 chromosomes.

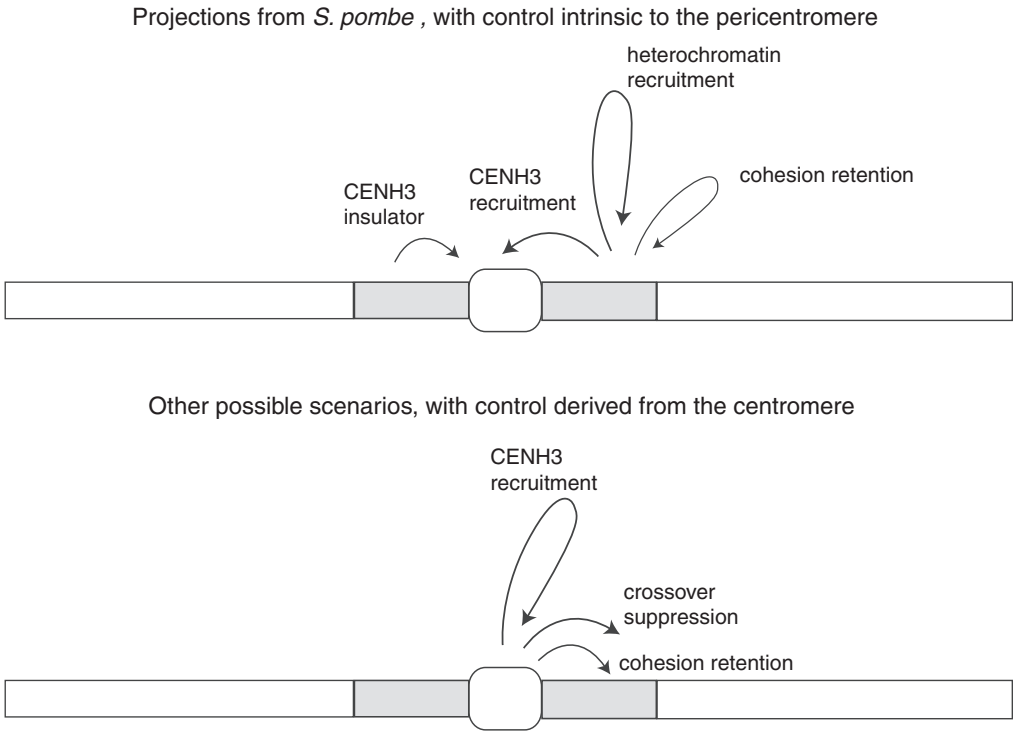

Figure 4. Models for centromere/pericentromere interactions. The arrows in each diagram depict proposed directions of causality, either arising from centromere (center circle) or pericentromere (flanking rectangles).

between CENH3 and the local chromatin environment (61,65–67).

Our findings reveal that the centromere is the single most prominent feature of the entire centromere-pericentromeric region in maize, and suggest that CENH3 alone may be responsible for functions often ascribed to pericentromeres (summarized in Figure 4). Cohesion may be controlled indirectly by kinetochores, and the vast domain of recombination suppression flanking centromeres may also be an indirect outcome of the total recombination block within centromere cores (68) and other mechanical constraints on recombination events that lie in close proximity to centromeres at meiosis [reviewed in (17)]. We emphasize, however, that our data do not demonstrate a mechanistic connection between the centromere and whole-chromosome patterns of RNA expression and chromatin marks. Furthermore, our data leave open the possibility that specific epigenetic marks such as phosphorylation or other transient modifications can be associated with pericentromeric areas during mitosis (6,7). What seems clear is that pericentromere structures vary dramatically across species but pericentromere function is conserved. Our data, combined with multiple studies pointing to similar conclusions, suggest that new models are needed to explain how and why kinetochores are invariably flanked by reduced recombination and regulated retention of cohesion.

ACCESSION NUMBER

NCBI Sequence Read Archive accession SRR218319.1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–3.

ACKNOWLEDGEMENTS

The authors thank Xiaoyu Zhang, Richard M. Gell, Gernot G. Presting, Jeffrey Ross-Ibarra, and Nathanael A. Ellis for expert advice and helpful discussions; Yeching Huang, Tao Zhang and Yufeng Wu for technical help with bioinformatic software; and Sarah Wolf, Rashin Ghaffari and Kathryn S. Hodges for logistical support.

FUNDING

National Science Foundation (0421671, 0607123, 0922703); National Institutes of Health (NIGMS F32GM095223). Funding for open access charge: National Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Black,B.E. and Cleveland,D.W. (2011) Epigenetic centromere propagation and the nature of CENP-a nucleosomes. *Cell*, **144**, 471–479.
2. Henikoff,S. and Furuyama,T. (2010) Epigenetic Inheritance of Centromeres. *Cold Spring Harb. Symp. Quant. Biol*, **75**, 51–60.
3. Nasmyth,K. and Haering,C.H. (2009) Cohesin: its roles and mechanisms. *Annu. Rev. Genet*., **43**, 525–558.

4. Bernard,P., Maure,J.F., Partridge,J.F., Genier,S., Javerzat,J.P. and Allshire,R.C. (2001) Requirement of heterochromatin for cohesion at centromeres. *Science*, **294**, 2539–2542.
5. Volpe,T., Kidner,C., Hall,I., Teng,G., Grewal,S. and Martienssen,R. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science*, **297**, 1833–1837.
6. Zhang,X., Li,X., Marshall,J.B., Zhong,C.X. and Dawe,R.K. (2005) Phosphoserines on maize CENTROMERIC HISTONE H3 and histone H3 demarcate the centromere and pericentromere during chromosome segregation. *Plant Cell*, **17**, 572–583.
7. Kawashima,S.A., Yamagishi,Y., Honda,T., Ishiguro,K. and Watanabe,Y. (2010) Phosphorylation of H2A by Bub1 prevents chromosomal instability through localizing shugoshin. *Science*, **327**, 172–177.
8. Folco,H., Pidoux,A., Urano,T. and Allshire,R. (2008) Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. *Science*, **319**, 94–97.
9. Kagansky,A., Folco,H., Almeida,R., Pidoux,A., Boukaba,A., Simmer,F., Urano,T., Hamilton,G. and Allshire,R. (2009) Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. *Science*, **324**, 1716–1719.
10. Chen,E.S., Zhang,K., Nicolas,E., Cam,H.P., Zofall,M. and Grewal,S.I. (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature*, **451**, 734–737.
11. Djupedal,I., Portoso,M., Spåhr,H., Bonilla,C., Gustafsson,C., Allshire,R. and Ekwall,K. (2005) RNA Pol II subunit Rpb7 promotes centromeric transcription and RNAi-directed chromatin silencing. *Genes Dev.*, **19**, 2301–2306.
12. Cam,H.P., Sugiyama,T., Chen,E.S., Chen,X., FitzGerald,P.C. and Grewal,S.I. (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat. Genet.*, **37**, 809–819.
13. Sullivan,B.A. and Karpen,G.H. (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat. Struct. Mol. Biol.*, **11**, 1076–1083.
14. Allis,D.C., Jenuwein,T., Reinberg,D. and Caparros,M.-L. (eds), (2007) *Epigenetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
15. Ekwall,K. (2007) Epigenetic control of centromere behavior. *Annu. Rev. Genet.*, **41**, 63–81.
16. Lu,J. and Gilbert,D.M. (2008) Cell cycle regulated transcription of heterochromatin in mammals vs. fission yeast: functional conservation or coincidence? *Cell Cycle*, **7**, 1907–1910.
17. Topp,C.N. and Dawe,R.K. (2006) Reinterpreting pericentromeric heterochromatin. *Curr. Opin. Plant Biol.*, **9**, 647–653.
18. Alonso,A., Hasson,D., Cheung,F. and Warburton,P.E. (2010) A paucity of heterochromatin at functional human neocentromeres. *Epigenetics Chromatin*, **3**, 6.
19. Smith,K.M., Phatale,P.A., Sullivan,C.M., Pomraning,K.R. and Freitag,M. (2011) Heterochromatin is required for normal distribution of Neurospora CenH3. *Mol. Cell. Biol*, **31**, 2528–2542.
20. Kang,J., Chaudhary,J., Dong,H., Kim,S., Brautigam,C.A. and Yu,H. (2011) Mitotic centromeric targeting of HP1 and its binding to Sgo1 are dispensable for sister-chromatid cohesion in human cells. *Mol. Biol. Cell*, **22**, 1181–1190.
21. Koch,B., Kueng,S., Ruckenbauer,C., Wendt,K.S. and Peters,J.M. (2008) The Suv39h-HP1 histone methylation pathway is dispensable for enrichment and protection of cohesin at centromeres in mammalian cells. *Chromosoma*, **117**, 199–210.
22. Serrano,A., Rodríguez-Corsino,M. and Losada,A. (2009) Heterochromatin protein 1 (HP1) proteins do not drive pericentromeric cohesin enrichment in human cells. *PLoS One*, **4**, e5118.
23. Turck,F., Roudier,F., Farrona,S., Martin-Magniette,M.L., Guillaume,E., Buisine,N., Gagnot,S., Martienssen,R.A., Coupland,G. and Colot,V. (2007) Arabidopsis TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet.*, **3**, e86.
24. Xu,L. and Shen,W.H. (2008) Polycomb silencing of KNOX genes confines shoot stem cell niches in Arabidopsis. *Curr. Biol.*, **18**, 1966–1971.
25. Zhang,X., Germann,S., Blus,B.J., Khorasanizadeh,S., Gaudin,V. and Jacobsen,S.E. (2007) The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.*, **14**, 869–871.
26. Schnable,P., Ware,D., Fulton,R., Stein,J., Wei,F., Pasternak,S., Liang,C., Zhang,J., Fulton,L., Graves,T. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.
27. Wolfgruber,T., Sharma,A., Schneider,K., Albert,P., Koo,D., Shi,J., Gao,Z., Han,F., Lee,H., Xu,R. *et al.* (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. *PLoS Genet.*, **5**, e1000743.
28. Jin,W., Lamb,J.C., Zhang,W., Kolano,B., Birchler,J.A. and Jiang,J. (2008) Histone modifications associated with both A and B chromosomes of maize. *Chromosome Res.*, **16**, 1203–1214.
29. Shi,J. and Dawe,R.K. (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics*, **173**, 1571–1583.
30. Houben,A., Demidov,D., Gernand,D., Meister,A., Leach,C.R. and Schubert,I. (2003) Methylation of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA content. *Plant J.*, **33**, 967–973.
31. Gent,J.I., Schvarzstein,M., Villeneuve,A.M., Gu,S.G., Jantsch,V., Fire,A.Z. and Baudrimont,A. (2009) A Caenorhabditis elegans RNA-directed RNA polymerase in sperm development and endogenous RNA interference. *Genetics*, **183**, 1297–1314.
32. Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science*, **294**, 858–862.
33. Pfeffer,S., Lagos-Quintana,M. and Tuschl,T. (2005) Cloning of small RNA molecules. *Curr. Protoc. Mol. Biol.*, **Chapter 26**, Unit 26.24.
34. Nagaki,K., Talbert,P., Zhong,C., Dawe,R., Henikoff,S. and Jiang,J. (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. *Genetics*, **163**, 1221–1225.
35. Zhong,C., Marshall,J., Topp,C., Mroczek,R., Kato,A., Nagaki,K., Birchler,J., Jiang,J. and Dawe,R. (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell*, **14**, 2825–2836.
36. Lister,R., Pelizzola,M., Dowen,R.H., Hawkins,R.D., Hon,G., Tonti-Filippini,J., Nery,J.R., Lee,L., Ye,Z., Ngo,Q.M. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, **462**, 315–322.
37. Griffiths-Jones,S. (2004) The microRNA Registry. *Nucleic Acids Res.*, **32**, D109–D111.
38. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
39. Tenaillon,M.I., Hufford,M.B., Gaut,B.S. and Ross-Ibarra,J. (2011) Genome size and transposable element content as determined by high-throughput sequencing in maize and Zea luxurians. *Genome Biol. Evol.*, **3**, 219–229.
40. Chen,P.Y., Cokus,S.J. and Pellegrini,M. (2010) BS Seeker: precise mapping for bisulfite sequencing. *BMC Bioinformatics*, **11**, 203.
41. Jia,Y., Lisch,D., Ohtsu,K., Scanlon,M., Nettleton,D. and Schnable,P. (2009) Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PLoS Genet.*, **5**, e1000737.
42. Nobuta,K., Lu,C., Shrivastava,R., Pillay,M., De Paoli,E., Accerbi,M., Arteaga-Vazquez,M., Sidorenko,L., Jeong,D.H., Yen,Y. *et al.* (2008) Distinct size distribution of endogeneous siRNAs in maize: Evidence from deep sequencing in the mop1-1 mutant. *Proc. Natl Acad. Sci. USA*, **105**, 14958–14963.
43. Yan,H., Ito,H., Nobuta,K., Ouyang,S., Jin,W., Tian,S., Lu,C., Venu,R.C., Wang,G.L., Green,P.J. *et al.* (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. *Plant Cell*, **18**, 2123–2133.

44. Yan,H., Talbert,P., Lee,H., Jett,J., Henikoff,S., Chen,F. and Jiang,J. (2008) Intergenic locations of rice centromeric chromatin. *PLoS Biol.*, **6**, e286.
45. Gore,M.A., Chia,J.M., Elshire,R.J., Sun,Q., Ersoz,E.S., Hurwitz,B.L., Peiffer,J.A., McMullen,M.D., Grills,G.S., Ross-Ibarra,J. *et al.* (2009) A first-generation haplotype map of maize. *Science*, **326**, 1115–1117.
46. Liu,S., Yeh,C., Ji,T., Ying,K., Wu,H., Tang,H., Fu,Y., Nettleton,D. and Schnable,P. (2009) Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet.*, **5**, e1000733.
47. Law,J.A. and Jacobsen,S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.*, **11**, 204–220.
48. Zhang,W., Lee,H., Koo,D. and Jiang,J. (2008) Epigenetic modification of centromeric chromatin: hypomethylation of DNA sequences in the CENH3-associated chromatin in Arabidopsis thaliana and maize. *Plant Cell*, **20**, 25–34.
49. Yan,H., Kikuchi,S., Neumann,P., Zhang,W., Wu,Y., Chen,F. and Jiang,J. (2010) Genome-wide mapping of cytosine methylation revealed dynamic DNA methylation patterns associated with genes and centromeres in rice. *Plant J*, **636**, 353–365.
50. Koo,D.H., Han,F., Birchler,J.A. and Jiang,J. (2011) Distinct DNA methylation patterns associated with active and inactive centromeres of the maize B chromosome. *Genome Res*, **21**, 908–914.
51. Koo,D.H. and Jiang,J. (2009) Super-stretched pachytene chromosomes for fluorescence in situ hybridization mapping and immunodetection of DNA methylation. *Plant J.*, **59**, 509–516.
52. Feng,S., Cokus,S.J., Zhang,X., Chen,P.Y., Bostick,M., Goll,M.G., Hetzel,J., Jain,J., Strauss,S.H., Halpern,M.E. *et al.* (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl Acad. Sci. USA*, **107**, 8689–8694.
53. Zemach,A., McDaniel,I.E., Silva,P. and Zilberman,D. (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, **328**, 916–919.
54. Jin,W., Melo,J.R., Nagaki,K., Talbert,P.B., Henikoff,S., Dawe,R.K. and Jiang,J. (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. *Plant Cell*, **16**, 571–581.
55. Matzke,M., Kanno,T., Daxinger,L., Huettel,B. and Matzke,A.J. (2009) RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.*, **21**, 367–376.
56. Topp,C., Zhong,C. and Dawe,R. (2004) Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc. Natl Acad. Sci. USA*, **101**, 15986–15991.
57. Vourc'h,C. and Biamonti,G. (2011) Transcription of Satellite DNAs in Mammals. *Prog. Mol. Subcell. Biol.*, **51**, 95–118.
58. Wang,X., Elling,A.A., Li,X., Li,N., Peng,Z., He,G., Sun,H., Qi,Y., Liu,X.S. and Deng,X.W. (2009) Genome-wide and organ-specific landscapes of epigenetic modifications and their relationships to mRNA and small RNA transcriptomes in maize. *Plant Cell*, **21**, 1053–1069.
59. Du,Y., Topp,C. and Dawe,R. (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet.*, **6**, e1000835.
60. Singh,M., Goel,S., Meeley,R.B., Dantec,C., Parrinello,H., Michaud,C., Leblanc,O. and Grimanelli,D. (2011) Production of viable gametes without meiosis in maize deficient for an ARGONAUTE protein. *Plant Cell*, **23**, 443–458.
61. Bergmann,J.H., Rodríguez,M.G., Martins,N.M., Kimura,H., Kelly,D.A., Masumoto,H., Larionov,V., Jansen,L.E. and Earnshaw,W.C. (2010) Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J*, **30**, 328–340.
62. Ribeiro,S.A., Vagnarelli,P., Dong,Y., Hori,T., McEwen,B.F., Fukagawa,T., Flors,C. and Earnshaw,W.C. (2010) A super-resolution map of the vertebrate kinetochore. *Proc. Natl Acad. Sci. USA*, **107**, 10484–10489.
63. Weber,S.A., Gerton,J.L., Polancic,J.E., DeRisi,J.L., Koshland,D. and Megee,P.C. (2004) The kinetochore is an enhancer of pericentric cohesin binding. *PLoS Biol.*, **2**, E260.
64. Han,F., Lamb,J.C. and Birchler,J.A. (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc. Natl Acad. Sci. USA*, **103**, 3238–3243.
65. Nakashima,H., Nakano,M., Ohnishi,R., Hiraoka,Y., Kaneda,Y., Sugino,A. and Masumoto,H. (2005) Assembly of additional heterochromatin distinct from centromere-kinetochore chromatin is required for de novo formation of human artificial chromosome. *J. Cell Sci.*, **118**, 5885–5898.
66. Lam,A.L., Boivin,C.D., Bonney,C.F., Rudd,M.K. and Sullivan,B.A. (2006) Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc. Natl Acad. Sci. USA*, **103**, 4186–4191.
67. Okamoto,Y., Nakano,M., Ohzeki,J., Larionov,V. and Masumoto,H. (2007) A minimal CENP-A core is required for nucleation and maintenance of a functional human centromere. *EMBO J.*, **26**, 1279–1291.
68. Shi,J., Wolf,S.E., Burke,J.M., Presting,G.G., Ross-Ibarra,J. and Dawe,R.K. (2010) Widespread gene conversion in centromere cores. *PLoS Biol.*, **8**, e1000327.