Species-specific satellite DNA composition dictates *de novo* formation of PRC1-mediated pericentric heterochromatin

Piero Lamelza[1], Malena Parrado[1], Emma Hamlin[1], Michael A. Lampson[1]

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA, 19104

Address correspondence to: plamelza@gmail.com or lampson@sas.upenn.edu

Abstract

Pericentromeres are heterochromatic regions adjacent to centromeres that ensure accurate chromosome segregation. Despite their conserved function, they are composed of rapidly evolving A/T-rich satellite DNA. This paradoxical observation is partially resolved by epigenetic mechanisms that maintain heterochromatin, independent of specific DNA sequences. However, it is unclear how satellite DNA sequence variation impacts *de novo* formation of pericentric heterochromatin, which is initially absent from paternal chromosomes in the zygote. Here we show that satellite variation has functional consequences for zygotic heterochromatin formation, recruitment of the Chromosome Passenger Complex (CPC), and interactions with spindle microtubules. In *M. musculus* zygotes, Polycomb Repressive Complex 1 (PRC1) is recruited to pericentric satellites by its AT-hook domain, which binds runs of A/T nucleotides, to generate H2AK119ub1 (H2Aub) heterochromatin. By fertilizing *M. musculus* eggs with sperm from other mouse species, we show that species-specific satellite sequences differ in their ability to recruit PRC1 and form H2Aub. This satellite-DNA mediated increase in PRC1 heterochromatin leads to reduced CPC recruitment and increased microtubule forces on kinetochores. Our results provide a direct link between satellite DNA composition and pericentromere function in the zygote, when epigenetic pathways maintaining pericentromere heterochromatin are absent.

Introduction

Satellite DNA is composed of tandemly repeated DNA sequences that can make up some of the largest non-coding regions of eukaryotic genomes. Satellites are among the most rapidly evolving genomic sequences, with their monomer size, nucleotide sequence, genomic distribution and abundance varying widely between species[1,2] . Despite these rapid changes, they often underly

chromosomal loci with highly conserved functions[3,4]. Recent advances in long-read sequencing technology have made huge progress in precisely characterizing the genetic composition and diversity of repetitive satellite arrays[5–7]. However, it is still largely unknown what functions satellite DNA may have at these essential loci and even less clear what functional consequences may arise from the natural diversity of these sequences.

Centromeres and pericentromeres, two loci with conserved functions in orchestrating accurate chromosome segregation during cell division, are composed of rapidly evolving satellite arrays in many species. This paradoxical observation is partially resolved by evidence indicating that centromere and pericentromere function, packaging and inheritance rely on epigenetic mechanisms. The histone H3 variant CENP-A (or CenH3) defines centromeres in most eukaryotes[8], while pericentric satellites are enriched with H3K9me3-based constitutive heterochromatin[9]. Both CENP-A and H3K9me3 serve as epigenetic templates for their inheritance through mitotic cell divisions via self-propagating epigenetic loops. CENP-A recruits the machinery for assembling new CENP-A nucleosomes[10], and H3K9me3 recruits histone methyltransferases that catalyze formation of new H3K9me3[11]. Because these epigenetic mechanisms operate independently of DNA sequence, variation in satellite DNA composition was thought to have little or no functional impact. However, during spermiogenesis in various metazoans, most histones are removed from the genome and replaced by protamines[12,13]. After fertilization, maternal machinery promptly removes protamines from the paternal genome and repackages it in maternally supplied histones[14]. CENP-A is the exception to this rule, as it is retained though spermiogenesis and thus serves as the epigenetic template for centromere inheritance through the male germline[15]. The absence of H3K9me3 and other potential epigenetic templates from paternal chromatin, however, suggests that pericentric satellite DNA may provide a genetic pathway to direct the *de novo* formation of pericentric heterochromatin in the zygote. Furthermore, this potential genetic pathway implies that variation in satellite DNA may have functional consequences for pericentromere function at this crucial stage of development.

In the widely used mammalian model species *Mus musculus* (house mouse), it takes approximately four cell cycles after fertilization to fully restore H3K9me3 heterochromatin at paternal pericentromeres. In the meantime, histone modifications catalyzed by Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2) package paternal pericentromeres in what is typically considered facultative heterochromatin (H2AK119ub1 and H3K27me3, respectively)[16–18]. Both modifications contribute to the transcriptional silencing of the 240 bp A/T-rich “major satellite” repeats at *M. musculus* pericentromeres[16,19], with PRC1 also crucial for maintaining the integrity of these arrays during the first embryonic mitosis[20]. Moreover, major satellite sequences can direct formation of PRC1 heterochromatin. PRC1 localization to paternal pericentromeres immediately after fertilization

relies on the AT-hook of one of its subunits, Cbx2[18]. The AT-hook is a five amino acid motif that has nanomolar binding affinity for the relatively narrow minor grooves of dsDNA generated by stretches of tandem A/T nucleotides (A/T runs)[21–25] (Fig 1A). It has therefore been proposed that PRC1 localization to paternal *M. musculus* pericentromeres is mediated by the AT-hook of Cbx2 directly binding the narrow minor grooves of the A/T runs present in major satellite repeats[18].

Results

Variation in PRC1 heterochromatin formed on divergent pericentric satellites

If major satellite DNA directs *de novo* PRC1 heterochromatin formation, we hypothesized that divergent pericentric satellite sequences would alter PRC1 binding and H2AK119ub1 formation. To compare paternal chromosomes with different pericentric sequences, we developed an experimental system based on fertilizing *M. musculus* eggs with sperm from one of three mouse species harboring genetically divergent pericentric satellites: *M. musculus, M. caroli* or *M. pahari* (Fig 1B). The maternal cytoplasm is constant in each case, so any differences in PRC1 heterochromatin established at paternal pericentromeres in the zygote reflect differences in paternal satellite sequences. We injected sperm heads into MII oocytes (i.e. intracytoplasmic injection, ICSI[26]) to bypass reproductive barriers between species, arrested the resulting zygotes at the first mitosis by kinesin-5 inhibition, and measured H2AK119ub1 by immunofluorescence. Because only maternal chromosomes have H3K9me3 at this stage, we used this mark to distinguish maternal and paternal chromosomes[16]. We found H2AK119ub1 enriched on paternal *M. musculus* and *M. caroli* pericentromeres relative to chromosome arms, but not on *M. pahari* pericentromeres (Fig 1C and 1F). In certain contexts, PRC1 is recruited to chromatin by the chromodomain of its Cbx subunit, which directly binds H3K27me3[27–29]. Since Cbx2 also contains a chromodomain (Fig 1A), we tested whether the absence of H2AK119ub1 at *M. pahari* paternal pericentromeres might be due to a lack of pericentric H3K27me3. In contrast, we found that all three species' pericentromeres robustly acquire H3K27me3 (Supp Fig 1A). This result indicates that H3K27me3 is not sufficient for pericentric H2AK119ub1 formation and supports previous findings that PRC2 and H3K27me3 are not required for PRC1 localization to paternal pericentromeres[16].

Satellite sequence determinants of PRC1 binding

To explain the differences in H2AK119ub1, we hypothesized that *M. musculus* and *M. caroli* pericentromeric satellites, but not *M. pahari* satellites, are enriched for narrow DNA minor grooves that bind the AT-hook of Cbx2 to recruit the PRC1 complex. To test this possibility, we used DAPI staining to quantify the amount of narrow minor grooves generated by A/T-runs at each species'

pericentromere, as DAPI specifically binds to these regions on the dsDNA molecule[30–32]. We find increased DAPI staining at *M. musculus* and *M. caroli* pericentromeres relative to chromosome arms, but not at *M. pahari* pericentromeres (Figure 1D, 1F and Supp Fig 1B). To exclude the possibility that DAPI enrichment at pericentromeres represents a relatively more compact chromatin state compared to the rest of chromosome, we stained with Sytox Green, an intercalating DNA-dye with no considerable DNA sequence preference[33]. None of the three species show enrichment of Sytox Green at their pericentromeres (Fig 1E, 1F and Supp Fig 1C). Furthermore, all three species exhibit similar intensities of Sytox Green staining at pericentromeres, suggesting similar states of chromatin compaction. Together, these results demonstrate that *M. musculus* and *M. caroli* pericentromeres are enriched for narrow A/T-rich minor grooves that can robustly bind the AT-hook of Cbx2, whereas *M. pahari* pericentromeres are not.

We considered two possibilities for the underlying differences in pericentric DNA sequence between species, leading to differences in DAPI enrichment and acquisition of H2AK119ub1. We focused on comparisons between *M. musculus* and *M. pahari* because of the long-read sequencing data available for these species. First, *M. musculus* major satellites may have a higher frequency of A/T runs compared to *M. pahari* pericentric satellites (pi satellites). However, major and pi satellite consensus sequences have similar frequencies of A/T runs of lengths ranging from four (4W) to ten or more ($\underline{\geq}$10W) (Supp Fig 2A). To capture nucleotide diversity that may be lost in consensus sequences, we also calculated the frequencies of A/T runs across genomic contigs containing either major or pi satellite arrays. Consistent with consensus sequences, we find that both species' satellite arrays have similar frequencies of A/T runs (~825 vs ~750 per 10 kb window in major satellite or pi satellite, respectively) (Fig 2A). The frequencies are also similar for each length of A/T runs (Fig 2A and Supp Fig 2B), with the exceptions of 6W and $\underline{\geq}$10W, which are relatively more abundant in major and pi satellite arrays, respectively. However, it is unclear how differences in these lengths specifically would lead to higher PRC1 at *M. musculus* major satellite arrays.

The second possibility is that *M. musculus* major satellite arrays might be enriched for specific A/T sequences optimal for Cbx2 AT-hook binding, compared to *M. pahari* satellite arrays. We analyzed the sequence composition of 4W, 5W and 6W A/T runs which make up ~75% of the A/T stretches in both species' satellite arrays and therefore most potential AT-hook binding sites. The vast majority of 4W sequences in *M. musculus* major satellite arrays are AAAT (114 per 10 kb), AAAA (113 per 10 kb) or TAAA (39 per 10 kb) (Fig 2B and Supp Fig 2C). *M. pahari* pi satellite arrays have a similar frequency of TAAA sites (37 per 10 kb) but much lower frequencies of AAAT and AAAA sites (7.2 and 46 per 10 kb, respectively). Based on these differences, we hypothesized that the Cbx2 AT-hook preferentially binds AAAT and/or AAAA.

To test the binding preferences of the Cbx2 AT-hook, we used AlphaFold3[34] to perform *in silico* pairwise competitive binding assays for AAAT and AAAA vs all other possible A/T tetranucleotides (Fig 2C). For each competition assay, we model a single dsDNA molecule encoding two distinct A/T runs flanked by four G:C base pairs (5' GGGGWWWWGGGGWWWWGGGG 3'). These two A/T runs "compete" to bind a single peptide consisting of the core Cbx2 AT-hook along with flanking amino acids (RKRGKR**PRGRP**RKHTVTSS). All AlphaFold models predict that the Cbx2 AT-hook prefers to bind AAAT and AAAA over all other A/T tetranucleotides (Fig 2C). When AAAT and AAAA compete against each other, their relative order on the dsDNA dictates which sequence binds the AT-hook (Supp Fig 2D, see Materials and Methods), indicating that these two sequences have similar affinity for the AT-hook.

We obtain similar results when analyzing 5W and 6W A/T runs. AAAAT and AAAAAT are the most abundant 5W and 6W sequences in the *M. musculus* major satellite array (147 per 10 kb and 70 per 10 kb, respectively) (Fig 2B and Supp Fig 2C). These sequences outcompete all other prevalent 5W and 6W sequences in both major and pi satellite arrays for Cbx2 AT-hook binding (Fig 2C). Furthermore, AAAAT and AAAAAT are 8 and 2.5-fold less frequent, respectively, in *M. pahari* pi satellite compared to *M. musculus* major satellite arrays (Fig 2B and Supp Fig 2C). These results indicate that the Cbx2 AT-hook prefers to bind stretches of adenine nucleotides that end with a single thymine (A_nT), consistent with previously reports showing that the core motif of the AT-hook (PRGRP) preferentially binds the minor groove of the sequence 5' AA(A/T)T 3'. This preference reflects optimal Van der Waals packing when adenine bases are on opposite sides of the A/T run[25]. Additionally, a meta-analysis of PDB structures[35] indicates that this sequence generates one of the narrowest minor grooves, which may also contribute to preferential binding. To generalize this analysis to include A_nT sequences (like AAAT) present in longer A/T runs, we quantified the frequency of A_nT sequences of various lengths (i.e. AAAT, AAAAT, AAAAAT, AAAAAAT and AAAAAAAT) in major and pi satellite arrays independent of total A/T run length. We find that major satellite arrays have approximately three times more of these A_nT sequences than *M. pahari* pi satellite arrays (Fig 2D and Supp Fig 2E). Furthermore, pi satellite arrays exhibit A_nT frequencies similar to euchromatin. Overall, these analyses indicate that the abundant A_nT sequences in *M. musculus* major satellite recruit Cbx2 and the rest of the PRC1 complex to generate H2AK119ub1. In contrast, both *M. pahari* pi satellite arrays and euchromatin fail to enrich PRC1 and therefore largely lack H2AK119ub1.

Functional consequences of variation in PRC1 heterochromatin

Next, we asked if the difference in PRC1-heterochromatin between paternal *M. musculus* and *M. pahari* chromosomes has consequences for pericentromere function. Notably, H2AK119ub1 is

positioned close to the site of another post-translational modification: phosphorylation of H2AT121 by Bub1 kinase. H2AT121phos recruits the Chromosome Passenger Complex (CPC) to pericentromeres via Sgo1 (Fig 3A) to ensure accurate chromosome segregation by regulating kinetochore-microtubule interactions during mitosis[36–38]. ChIP-seq and *in vitro* assays on reconstituted nucleosomes have shown that H2AK119ub1 and H2AT121phos are mutually exclusive[39], suggesting that one modification physically blocks the deposition of the other on the same H2A C-terminal tail. Given that pericentric PRC1 heterochromatin forms immediately after fertilization[16,18,20,40], we hypothesized that H2AK119ub1 occludes Bub1 kinase from phosphorylating H2AT121 and thereby downregulates the CPC. This hypothesis predicts low H2AT121phos and low CPC on paternal *M. musculus* pericentromeres (Fig 3A), which have high H2AK119ub1 compared to maternal *M. musculus* and paternal *M. pahari* pericentromeres (Fig 1C, F). Furthermore, the Aurora B kinase subunit of the CPC phosphorylates kinetochore proteins to destabilize kinetochore-microtubule interactions[41]. In tissue culture cells, expression of a non-phosphorylatable mutant of a key Aurora B substrate, Hec1, results in increased stabilization of kinetochore-microtubule interactions and increased forces pulling the sister kinetochores apart at metaphase[42]. Increased inter-kinetochore distance therefore provides a functional readout for reduced Aurora B activity (Fig 3A).

To test these predictions, we first visualized H2AT121phos, the CPC subunit Survivin, and the PRC1 subunit Rnf2, which remains associated with paternal *M. musculus* pericentromeres during mitosis. In *M. musculus* x *M. musculus* zygotes arrested at metaphase, we found that H2AT121phos is present along the entire chromosome (Fig 3B) as reported in the previous cell cycle in mouse oocytes[43,44]. Moreover, we found an inverse relationship between Rnf2 and H2AT121phos, with the presence of Rnf2 at paternal pericentromeres coinciding with a reduction in H2AT121phos (Fig 3B). In contrast, maternal pericentromeres with undetectable Rnf2 show no reduction in H2AT121phos. On average, H2AT121phos is reduced to ~60% at paternal *M. musculus* pericentromeres compared to maternal pericentromeres (Fig 3B). Consistent with this result, Survivin is reduced to ~65% on paternal pericentromeres compared to maternal pericentromeres in *M. musculus* x *M. musculus* zygotes. In contrast, Survivin is similar between paternal and maternal pericentromeres in *M. musculus* x *M. pahari* zygotes (Fig 3C), consistent with the absence of PRC1 from all pericentromeres in this hybrid (Fig 1C). Together, these results show that PRC1 heterochromatin inhibits the acquisition of H2AT121phos and thereby reduces CPC recruitment to pericentromeres of paternal *M. musculus* chromosomes.

To test for functional consequences of CPC differences between pericentromeres in the zygote, we measured the distance between sister kinetochores at metaphase. We found increased inter-kinetochore distances for paternal *M. musculus* chromosomes compared to either maternal *M.*

musculus or paternal *M. pahari* chromosomes (Fig 3D). This result is consistent with our hypothesis that PRC1 heterochromatin inhibits H2AT121 phosphorylation, downregulates the CPC, and stabilizes kinetochore-microtubule interactions on paternal *M. musculus* chromosomes. Increased inter-kinetochore distances could theoretically also reflect reduced pericentric cohesion because in addition to recruiting the CPC, Sgo1 also recruits phosphatases that counteract cohesin removal that occurs along chromosome arms during mitotic prophase[45,46] (Fig 3A). However, sister chromatids appear in close proximity along most of their lengths during zygotic mitosis (Fig 3D)[47], suggesting that cohesin is not removed during prophase at this stage of development. Therefore, we interpret the increase in inter-kinetochore distance as a consequence of reduced CPC activity.

Discussion

Overall, our results show that divergent satellite DNA sequences, through differential recruitment of a heterochromatin forming complex, lead to functionally diverse pericentromeres. In contrast to the notion that AT-hooks have a high affinity but low specificity for A/T-rich minor grooves, our AlphaFold 3 modeling suggests that the AT-hook of Cbx2$^{M.\ musculus}$ prefers to bind A_nT sequences, which is consistent with the known optimal binding sequence of other AT-hook proteins[25]. *M. musculus* major satellite arrays are enriched for A_nT sequences compared to the rest of the genome, whereas *M. pahari* pi satellite arrays are not. This variation in satellite DNA composition is an example of the extensive diversity revealed by recent and ongoing long-read sequencing efforts. However, experimental systems that test functional impacts of this diversity remain limited. For example, known satellite variation within a mouse species is limited to copy-number variation of the same satellite[43,48,49]. Our hybrid embryo system is a powerful model to probe functional consequences of satellite DNA evolution incorporating sequence variation between satellites.

Despite the importance of PRC1 heterochromatin in *M. musculus* zygotes, our findings suggest that it is not a conserved feature of paternal pericentromeres during mouse preimplantation development, as it is lacking from *M. pahari* pericentromeres. *M. musculus* major satellites and *M. caroli* satellites may have specific properties that require their packaging in PRC1 heterochromatin, whereas *M. pahari* pi satellites can rely on PRC2-mediated H3K27me3. Alternatively, the *M. pahari* egg cytoplasm may have specific adaptations that allow PRC1 to be recruited to pericentromeres by a different mechanism. Because the core Cbx2 AT-hook sequence is conserved across Muridae (data not shown), it is unlikely that the AT-hook of Cbx2$^{M.\ pahari}$ has adapted to bind A/T sequences enriched at *M. pahari* pi satellites. Similarly, the Cbx2 chromodomain is identical across Muridae (data not shown), indicating that chromodomain of Cbx2$^{M.\ pahari}$ has not adapted to have a higher

binding affinity towards H3K27me3. However, there are multiple PRC1 complexes that share the core ubiquitin ligase subunits (RING1A/1B) but differ in associated proteins that diversify PRC1 function, including the mechanism of recruitment to chromatin[50]. Thus, another targeting subunit expressed in *M. pahari* zygotes might recognize paternal *M. pahari* pericentromeres.

Our findings highlight the first zygotic cell cycle as a unique environment where *de novo* heterochromatin formation and function are sensitive to variation in satellite DNA sequence, due to the loss of epigenetic templates from the paternal genome during spermiogenesis. This functional variation occurs at a critical stage of development where chromosome segregation errors would lead to aneuploidy of all subsequent cells. Despite the importance of this first cell division, PRC1 heterochromatin inhibits molecular pathways that ensure accurate chromosome segregation, particularly regulation of kinetochore-microtubule attachments by the CPC. Our findings suggest that expansion of pericentric satellite arrays enriched for A_nT sequences would lead to increased PRC1 heterochromatin and a corresponding decline in CPC recruitment, making paternal chromosomes more prone to segregation errors. Consequently, we propose that large major satellite arrays are selected against in natural populations. Consistent with this idea, wild-caught *M. musculus* have roughly ten times fewer major satellite repeats than inbred laboratory strains, where natural selection is relatively weak[48]. Despite potentially deleterious effects, however, pericentric PRC1 heterochromatin is required for the transcriptional silencing and stability of paternal major satellite arrays[18,20], suggesting an evolutionary trade-off between proper packaging of major satellites and chromosome segregation fidelity.

Materials and Methods

Mouse strains

The mouse strain representing *Mus musculus* in our experiments is FVB/NJ, purchased from Jackson Laboratory (strain#001800). FVB/NJ was chosen because its MII eggs can survive intracytoplasmic sperm injection and because it is an inbred strain. *M. caroli* (CAROLI/EiJ, strain#000926) and *M. pahari* (PAHARI/EiJ, strain #002655) were also purchased from Jackson Laboratory. We now maintain our own colony of *M. pahari* because Jackson Laboratory no longer carries this species. All animal experiments and protocols were approved by the Institutional Animal Use and Care Committee of the University of Pennsylvania and are consistent with National Institutes of Health guidelines (protocol #804882).

Intracytoplasmic sperm injection (ICSI)

To prepare sperm for ICSI, the epididymis and vas deferens were removed from males and dissected in PBS to release mature sperm. Sperm were allowed to swim-out for 10 mins on a 37°C slide warmer. After swim out, sperm were pelleted by centrifugation at 700xg at 4°C for 5 min. The PBS supernatant was removed, and sperm were then washed twice with ice-cold Nuclear Isolation Media (NIM, 123 mM KCl, 2.6mM NaCl, 7.8 mM Na_2PO_4, 1.4 mM KH_2PO_4, 3 mM EDTA, pH adjusted to 7.2 using 1M KOH) with 1% poly vinyl alcohol (PVA). After the second wash, sperm were resuspended in 100 µL of NIM 1% PVA and then sonicated using a Branson Sonic Bath Model 1210 for 15-20 second intervals until at most 30% of sperm had their heads detached from their tails. Sonicated sperm were washed twice with NIM 1% PVA. After the second wash, the sperm pellet was resuspended in a 1:1 solution of glycerol and NIM 1% PVA and placed at -20°C until the day of injection (the following day or 1 week later).

Females were super-ovulated by injection with 5 U PMSG (Peptides International) followed by injection with 5 U hCG (Sigma) 48 hours later. MII oocytes were collected from females 14-15 hours post-hCG injection and placed in M2 media (Sigma) with hyaluronidase (0.15 mg/mL) to remove cumulus cells. MII eggs were then washed through four drops of M2 media supplemented with 4mg/mL BSA (M2+BSA) and left on a 37°C plate warmer until injection. 10-20 µLs of sonicated sperm was diluted in 200 µL of NIM 1% PVA, gently vortexed, and then washed twice with NIM 1% PVA. After the last wash, the sperm pellet was resuspended in 100 µL of NIM 1% PVA and left on ice until injection.

Batches of 10 MII eggs were injected with sperm heads at a time in room temperature drops of M2+BSA. After injection, each batch was moved to a drop of M2+BSA on a 37°C plate warmer and allowed to recover for 1 hour. After recovery, each batch was washed through four drops of pre-equilibrated AKSOM (Millipore Sigma) and placed in at 37°C humidified incubator with 5% CO_2 until injections were complete. Depending on the experiment, injected eggs were then moved into a drop of AKSOM with either 10 µM STLC (Sigma) (to generate monopolar spindles during mitosis) or 5 µM proTAME (R&D Systems) (to arrest zygotes in metaphase) at least three hours after injection and were allowed to develop overnight in a 37°C incubator with 5% CO_2.

Fixing and staining

The morning after ICSI, zygotes arrested at mitosis were fixed in 37°C 2% paraformaldehyde in PBS for 20 minutes, washed through three pools of blocking solution (PBS containing 0.5% BSA, 0.01% Tween-20) and then left at 4°C overnight. The next morning cells were permeabilized in PBS containing 0.5% Triton X-100 (Sigma) for 15 minutes at room temperature, quickly washed through two pools of blocking solution and then allowed to block for 20 minutes at room temperature. For

H3K9me3 staining, the cells were treated with lambda-phosphatase (1600 U, NEB) for 1 hour at 37°C. Embryos were then quickly washed through two pools of blocking solution and then incubated with primary antibodies for 1 hour in a dark humidified chamber at room temperature (or overnight at 4°C for H2AT121phos staining). Afterwards, embryos underwent three 15-minute washes in blocking solution and then were incubated with secondary antibodies for 1 hour in a dark humidified chamber at room temperature. Cells then underwent three 15-minute washes in blocking solution and were mounted in Vectashield with DAPI (Vector) to stain chromosomes. For cells that were stained with Sytox Green (Invitrogen), the first wash after incubation with secondary antibodies included 1 uM Sytox Green and after the next two washes were mounted in Vectashield without DAPI. Primary antibodies used are rabbit anti-H2AK119ub1 (1:800, Cell Signaling, D27C4), rabbit anti-H3K27me3 (1:700, Cell Signaling, C36B11), mouse anti-H3K9me3 (1:200, Active Motif, 39285), rabbit anti-Survivin (1:500, Cell Signaling, 71G4B7), rabbit anti-H2AT120phos (1:2500, Active Motif, 39392), mouse anti-Rnf2/Ring1B (1:500, Active Motif, 39664) and rabbit anti-CENP-C[51] (1:500). Secondary antibodies used are donkey anti-rabbit Alexa Fluor 488, donkey anti-mouse Alexa Fluor 594, donkey anti-rabbit Alexa Fluor 594, and donkey anti-mouse Alex Fluor 647. All secondary antibodies are purchased from Invitrogen and used at a dilution of 1:500.

Microscopy

Some confocal images were collected as z-stacks with 0.5 µm intervals, using a microscope (DMI4000 B; Leica) equipped with a 63× 1.3 NA glycerol-immersion objective lens, an xy piezo Z stage (Applied Scientific Instrumentation), a spinning disk confocal scanner (Yokogawa Corporation of America), an electron multiplier charge-coupled device camera (ImageEM C9100–13; Hamamatsu Photonics), and either an LMM5 (Spectral Applied Research) or Versalase (Vortran Laser Technology) laser merge module, controlled by MetaMorph software (Molecular Devices, v7.10.3.294). Confocal images were also collected using a Leica TCS SP8 Four Channel Spectral Confocal System with a 63x objective lens. A z-stack of 0.3 µm intervals was used to collect images of embryos used to calculate sister-kinetochore distances. All samples in an experiment were imaged using the same laser settings.

Image quantification and statistical analysis

All image analysis was carried out using ImageJ/Fiji[52]. To measure signal intensities along single chromosomes, we selected chromosomes that were mostly in the XY plane and generated max intensities projections of them. Beginning from the pericentric end, we then used the segmented line tool to manually draw a linear ROI (roughly the width of the chromosome) along the

chromosome's length and measured the average signal intensity per unit distance along the ROI. Average background was then subtracted from these values. To more easily compare the distributions of DAPI and Sytox Green intensity, their values were normalized to the average intensity within each cell/embryo using custom R scripts.

To measure Survivin intensity at pericentromere ends, we also selected chromosomes that were mostly in the XY plane and generated maximum intensity projections of them. We then use the rectangular selection tool to generate square ROIs that encompassed pericentric ends of chromosomes and measured average intensity within the ROI. Average background was then subtracted from these values. The resulting values were then normalized to maternal chromosomes in each cell/embryo using custom R scripts.

To measure the distance between sister-centromeres, we uniquely labelled each centromere and extracted its 3D coordinates using 3D maxima finder and 3D Roi Manager (based on CENP-C signal). We then manually assigned sister-centromeres with the help of BigDataViewer to "rotate" the image stack in cases where the spindle pole axis was not in the XY plane and then used custom R scripts to calculate the distance between them.

P-values were determined using R. We performed non-parametric Kruskal-Wallis tests followed by Dunn's test and used the Bonferroni method to correct *P*-values for multiple testing. Plots were made in the ggplot2 package in R.

Satellite sequence analysis

M. musculus assemblies containing Major satellite arrays were found by BLASTing the Major satellite consensus sequence[48] on NCBI BLAST server. These assemblies are from a long-read *M. musculus* genomes generated by the Mouse Genome Project and Darwin Tree of Life Project (https://www.nhm.ac.uk/our-science/research/projects/darwin-tree-of-life.html#:~:text=The%20Darwin%20Tree%20of%20Life%20project%20aims%20to%20generate%20DNA,fungi%20within%20the%20British%20Isles.). The Pi satellite array containing contigs are from a published *M. pahari* long-read assembly[49]. These contigs were chosen because they represent relatively complete assemblies of functional *M. pahari* centromeres and pericentromeres based on CENP-A and H3K9me3 ChIP-seq. We chose to study Major and Pi satellite arrays because they are the most abundant pericentric satellites and because PRC1 is known to associate with Major satellite[16].

To quantify the frequency of A/T runs of various lengths along a contig, we used the matchPattern() function in the Biostrings package in R (https://bioconductor.org/packages/Biostrings). Specifically, we searched for A/T runs that were at least four nucleotides long (i.e. matched

"WWWW"), with overlapping matches eventually being merged into one longer A/T run by using the reduce() function. To analyze the composition of 4W and 5W A/T runs, we used the matchPattern() function to find all possible 4W and 5W sequences and then only kept matches that overlapped with 4W or 5W A/T run lengths, respectively. To analyze the composition of 6W A/T runs, we used the 6W A/T run coordinates to extract sequences within the satellite array using the Genomic Ranges package[53] in R. Sequences that are reverse complements of each other are redundant and therefore were merged into one. The average number of times a particular sequence occurred per 10 kb bin within an array was calculated by dividing the total number of occurrences of a particular sequence within the array by the length of the array and then multiplying by 10,000. We defined the range of the satellite arrays for each contig based on coordinates from satellite consensus sequence BLAST results. Finally A_nT sequences were also found along contigs using the matchPattern() function. To prevent over-counting of smaller AnT sequences (e.g. overlap of AAAT and AAAAT), only the longest of overlapping A_nT sequences were counted.

AlphaFold3 modeling

All sequences were modeled using the AlphaFold3 webserver (https://alphafoldserver.com/). To model a single dsDNA molecule encoding two distinct A/T runs flanked by four G:C base pairs, we entered a single forward and a single reverse strand (e.g. forward: 5'GGGGAAAAGGGGAAATGGGG 3', reverse: 5'CCCCATTTCCCCTTTTCCCC 3'). We also modeled a single peptide encoding the AT-hook of Cbx2 with flanking amino acids (RKRGKR**PRGRP**RKHTVTSS). Alphafold3 outputs five model structures (0-4), with 0 scored as the best and the fourth scored as the worst. Each model was visually inspected using Pymol (The Pymol Molecular Graphics System, Version 3.0 Schrodinger, LLC). All five models for each competition assay were consistent with one another (data not shown). To rule out the possibility that the relative position along the dsDNA molecule dictates where the AT-hook binds, we flipped the relative positions of the two A/T sequences along the dsDNA. For example, we modeled both 5'GGGGAAAAGGGGAAATGGGG 3' and 5'GGGGAAATGGGGAAAAGGGG3' sequences (Figure S2D).

Acknowledgements

We thank Dr. Colin Conine for training in intracytoplasmic sperm injection and Dr. Ben Black and Dr. Damian Dudka for thoughtful conversations on AT-hook proteins. We also thank Dr. Black for initially suggesting that pericentric H2AK119ub1 may be incompatible with H2AT121phos.

Author Contributions

Conceptualization: P. Lamelza and M. A. Lampson, Methodology: P. Lamelza, M. Parrado and E. Hamlin, Investigation: P. Lamelza, M. Parrado and E. Hamlin, Writing: P. Lamelza and M. A. Lampson, Funding Acquisition: P. Lamelza and M.A. Lampson, Resources: M.A. Lampson

Conflict of Interest

The authors declare no competing financial interests.

Funding

This work was supported by: The Helen Hay Whitney Foundation and National Institutes of Health grant 1K99GM152835 awarded to P. Lamelza and National Institutes of Health grant R35GM122475 to M. A. Lampson.

Figure Legends

Figure 1. Abundance of narrow DNA minor grooves predicts the formation of pericentric PRC1 heterochromatin.

A) Schematic showing PRC1 recruitment to DNA via the AT-hook of its Cbx2 subunit, which binds into the minor groove of A/T runs present in satellite DNA. **B)** Mouse zygotes are generated by injecting a *M. musculus* egg arrested in meiosis II with the sperm of mouse species with divergent pericentric satellite DNAs. **C)** Zygotes generated with sperm from the indicated species were arrested in mitosis, then fixed and stained for H2AK119ub1 (green), H3K9me3 (magenta) to mark maternal *M. musculus* chromosomes, and DAPI (gray). **D)** Zygotes generated with *M. musculus* or *M. pahari* sperm, stained for Survivin (CPC subunit, green) to mark pericentromeres, H3K9me3 (magenta) to mark maternal chromosomes, and DAPI (gray). Arrowheads in insets point to paternal pericentromeres as indicated by presence of Survivin and absence of H3K9me3. **E)** Zygotes generated with *M. musculus* or *M. pahari* sperm, stained for H3K27me3 (green) to mark paternal pericentromeres and Sytox Green (gray). Arrowheads in insets point to paternal pericentromeres as indicated by enrichment of H3K27me3. **F)** Graphs plot average H2AK119ub1 (n=103-164), DAPI (n=72-165), and Sytox Green (n=44-66) intensities along *M. musculus* (red), *M. caroli* (yellow) and *M. pahari* (blue) paternal chromosomes, starting from pericentric ends. S.E.M. is indicated by light band surrounding the mean line. Images are max intensity z-projections, scale bars 10 μm or 1 μm (insets).

Figure 2. *M. musculus* major satellite arrays are enriched for A_nT sequences that preferentially bind the Cbx2 AT-hook

A) Histograms plotting the number of various A/T run lengths per 10 kb bin along portions of two genomic contigs containing either *M. musculus* Major satellite (left) or *M. pahari* Pi satellite arrays (right), indicated by tandem arrows above plots. **B)** The average number per 10 kb bin of all possible 4W (left) and 5W (right) sequences within the Major and Pi satellites arrays shown in A. **C)** Schematic (left) shows *in silico* competition assays between different A/T run sequences for the binding of a single peptide encoding the AT-hook of Cbx2, using AlphaFold3. Results (right) show preferred AT-hook binding to A_nT over all other A/T runs. **D)** Histograms plotting the number of various A_nT sequences per 10 kb bin along the same genomic contigs as in A.

Figure 3. PRC1 pericentric heterochromatin inhibits CPC recruitment

A) Schematic showing how H2AK119ub1 can inhibit H2AT121phos and CPC recruitment, resulting in increased distance between metaphase sister kinetochores at paternal *M. musculus* pericentromeres (right) vs *M. pahari* pericentromeres (left). **B)** Pure *M. musculus* zygotes were arrested at metaphase, fixed and stained for Rnf2 (PRC1 subunit, magenta) and H2AT121phos (green). Top and bottom insets show paternal (Rnf2 positive) and maternal (Rnf2 negative) chromosomes, respectively; arrowheads point to pericentromeric ends of chromosomes. Graphs plot average H2AT121phos and Rnf2 intensity along maternal (n=37) and paternal (n=37) chromosomes, starting from pericentric ends. **C)** Zygotes generated with *M. musculus* or *M. pahari* sperm, stained for Survivin (green) and H3K9me3 (magenta). Arrowheads in insets point to pericentric Survivin staining. Graphs plot the mean Survivin intensity of maternal and paternal pericentromeres in both types of zygotes. Each point represents a single pericentromere (n = 253-374 for each group) and boxes represent interquartile ranges. **D)** Zygotes generated with *M. musculus* or *M. pahari* sperm, stained for CENP-C (green) and H3K9me3 (magenta). Arrowhead points to a pair of paternal *M. musculus* sister kinetochores. Graphs plot distances between sister kinetochores for maternal and paternal chromosomes in both types of zygotes. Each point represents a single pair (n =230-284 for each group). All *P*-values were calculated by a Kruskal-Wallis test followed by Dunn's test using Bonferroni correction (*** $P < 0.001$). Images are max intensity z-projections; scale bars 10 μm or 1 μm (insets).

References

1. Thakur, J., Packiaraj, J. & Henikoff, S. Sequence, Chromatin and Evolution of Satellite DNA. *IJMS* **22**, 4309 (2021).

2. Melters, D. P. *et al.* Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**, R10 (2013).
3. Henikoff, S. The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science* **293**, 1098–1102 (2001).
4. Saint-Leandre, B. & Levine, M. T. The Telomere Paradox: Stable Genome Preservation with Rapidly Evolving Proteins. *Trends in Genetics* **36**, 232–242 (2020).
5. Nurk, S. *et al.* The complete sequence of a human genome. (2022).
6. Altemose, N. *et al.* Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178 (2022).
7. Logsdon, G. A. & Eichler, E. E. The Dynamic Structure and Rapid Evolution of Human Centromeric Satellite DNA. *Genes* **14**, 92 (2022).
8. Black, B. E. & Cleveland, D. W. Epigenetic Centromere Propagation and the Nature of CENP-A Nucleosomes. *Cell* **144**, 471–479 (2011).
9. Janssen, A., Colmenares, S. U. & Karpen, G. H. Heterochromatin: Guardian of the Genome. *Annu. Rev. Cell Dev. Biol.* **34**, 265–288 (2018).
10. French, B. T., Westhorpe, F. G., Limouse, C. & Straight, A. F. Xenopus laevis M18BP1 Directly Binds Existing CENP-A Nucleosomes to Promote Centromeric Chromatin Assembly. *Developmental Cell* **42**, 190-199.e10 (2017).
11. Fioriniello, S., Marano, D., Fiorillo, F., D'Esposito, M. & Della Ragione, F. Epigenetic Factors that Control Pericentric Heterochromatin Organization in Mammals. *Genes* **11**, 595 (2020).
12. Bao, J. & Bedford, M. T. Epigenetic regulation of the histone-to-protamine transition during spermiogenesis. *Reproduction* **151**, R55–R70 (2016).
13. Van Der Heijden, G. W. *et al.* Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. *Developmental Biology* **298**, 458–469 (2006).
14. Lin, C.-J., Koh, F. M., Wong, P., Conti, M. & Ramalho-Santos, M. Hira-Mediated H3.3 Incorporation Is Required for DNA Replication and Ribosomal RNA Transcription in the Mouse Zygote. *Developmental Cell* **30**, 268–279 (2014).
15. Das, A., Smoak, E. M., Linares-Saldana, R., Lampson, M. A. & Black, B. E. Centromere inheritance through the germline. *Chromosoma* **126**, 595–604 (2017).
16. Puschendorf, M. *et al.* PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet* **40**, 411–420 (2008).
17. Saksouk, N. *et al.* Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. *Molecular Cell* **56**, 580–594 (2014).

18. Tardat, M. *et al.* Cbx2 Targets PRC1 to Constitutive Heterochromatin in Mouse Zygotes in a Parent-of-Origin-Dependent Manner. *Molecular Cell* **58**, 157–171 (2015).

19. Santenard, A. *et al.* Heterochromatin formation in the mouse embryo requires critical residues of the histone variant H3.3. *Nat Cell Biol* **12**, 853–862 (2010).

20. Liu, Z. *et al.* SUMO ylated PRC 1 controls histone H3.3 deposition and genome integrity of embryonic heterochromatin. *EMBO J* **39**, (2020).

21. Kawaguchi, T., Machida, S., Kurumizaka, H., Tagami, H. & Nakayama, J. Phosphorylation of CBX2 controls its nucleosome-binding specificity. *The Journal of Biochemistry* **162**, 343–355 (2017).

22. Battista, S. *et al.* Binding to the Other Side: The AT-Hook DNA-Binding Domain Allows Nuclear Factors to Exploit the DNA Minor Groove. *IJMS* **25**, 8863 (2024).

23. Huth, J. R. *et al.* The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nature Structural Biology* **4,** 657-665 (1997).

24. Reeves, R. & Nissen, M. S. The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure. *Journal of Biological Chemistry* **265**, 8573–8582 (1990).

25. Reeves, R. Structure and Function of the HMGI(Y) Family of Architectural Transcription Factors. *Environmental health perspectives* **108**, 803-809 (2000).

26. Stein, P. & Schultz, R. M. ICSI in the Mouse. in *Methods in Enzymology* vol. 476 251–262 (Elsevier, 2010).

27. Wang, L. *et al.* Hierarchical Recruitment of Polycomb Group Silencing Complexes. *Molecular Cell* **14**, 637–646 (2004).

28. Cao, R. *et al.* Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298**, 1039–1043 (2002).

29. Blackledge, N. P., Rose, N. R. & Klose, R. J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643–649 (2015).

30. Breusegem, S. Y., Clegg, R. M. & Loontiens, F. G. Base-sequence specificity of Hoechst 33258 and DAPI binding to five (A/T)4 DNA sites with kinetic evidence for more than one high-affinity Hoechst 33258-AATT complex. *Journal of Molecular Biology* **315**, 1049–1061 (2002).

31. Tanious, F. A., Veal, J. M., Buczak, H., Ratmeyer, L. S. & Wilson, W. D. DAPI (4',6-diamidino-2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and intercalation at AU sites. *Biochemistry* **81**, 3103-3112 (1992).

32. Ren, J. & Chaires, J. B. Sequence and Structural Selectivity of Nucleic Acid Binding Ligands. *Biochemistry* **38**, 16067–16075 (1999).

33. Thakur, S., Cattoni, D. I. & Nöllmann, M. The fluorescence properties and binding mechanism of SYTOX green, a bright, low photo-damage DNA intercalating agent. *Eur Biophys J* **44**, 337–348 (2015).

34. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).

35. Rohs, R. *et al.* The role of DNA shape in protein–DNA recognition. *Nature* **461**, 1248–1253 (2009).

36. Carmena, M. The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nature Reviews. Molecular Cell Biology* **13,** 789-803 (2012).

37. Yamagishi, Y., Honda, T., Tanno, Y. & Watanabe, Y. Two Histone Marks Establish the Inner Centromere and Chromosome Bi-Orientation. *Science* **330**, 239–243 (2010).

38. Kawashima, S. A. *et al.* Shugoshin enables tension-generating attachment of kinetochores by loading Aurora to centromeres. *Genes Dev.* **21**, 420–435 (2007).

39. Aihara, H. *et al.* Histone H2A T120 Phosphorylation Promotes Oncogenic Transformation via Upregulation of Cyclin D1. *Molecular Cell* **64**, 176–188 (2016).

40. Eid, A. & Torres-Padilla, M.-E. Characterization of non-canonical Polycomb Repressive Complex 1 subunits during early mouse embryogenesis. *Epigenetics* **11**, 389–397 (2016).

41. McVey, S. L., Cosby, J. K. & Nannas, N. J. Aurora B Tension Sensing Mechanisms in the Kinetochore Ensure Accurate Chromosome Segregation. *IJMS* **22**, 8818 (2021).

42. DeLuca, J. G. *et al.* Kinetochore Microtubule Dynamics and Attachment Stability Are Regulated by Hec1. *Cell* **127**, 969–982 (2006).

43. Akera, T., Trimm, E. & Lampson, M. A. Molecular Strategies of Meiotic Cheating by Selfish Centromeres. *Cell* **178**, 1132-1144.e10 (2019).

44. Schober, C. S., Aydiner, F., Booth, C. J., Seli, E. & Reinke, V. The kinase VRK1 is required for normal meiotic progression in mammalian oogenesis. *Mechanisms of Development* **128**, 178–190 (2011).

45. Kitajima, T. S. *et al.* Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441**, 46–52 (2006).

46. Riedel, C. G. *et al.* Protein phosphatase 2A protects centromeric sister chromatid cohesion during meiosis I. *Nature* **441**, 53–61 (2006).

47. Yueh, W.-T., Singh, V. P. & Gerton, J. L. Maternal *Smc3* protects the integrity of the zygotic genome through DNA replication and mitosis. *Development* **148**, dev199800 (2021).

48. Arora, U. P., Charlebois, C., Lawal, R. A. & Dumont, B. L. Population and subspecies diversity at mouse centromere satellites. *BMC Genomics* **22**, 279 (2021).

49. Gambogi, C. W. *et al.* Centromere innovations within a mouse species. *SCIENCE ADVANCES* (2023).

50. Geng, Z. & Gao, Z. Mammalian PRC1 Complexes: Compositional Complexity and Diverse Molecular Mechanisms. *IJMS* **21**, 8594 (2020).

51. Das, A. *et al.* Epigenetic, genetic and maternal effects enable stable centromere inheritance. *Nat Cell Biol* **24**, 748–756 (2022).

52. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–682 (2012).

53. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* **9**, e1003118 (2013).

Figure 1

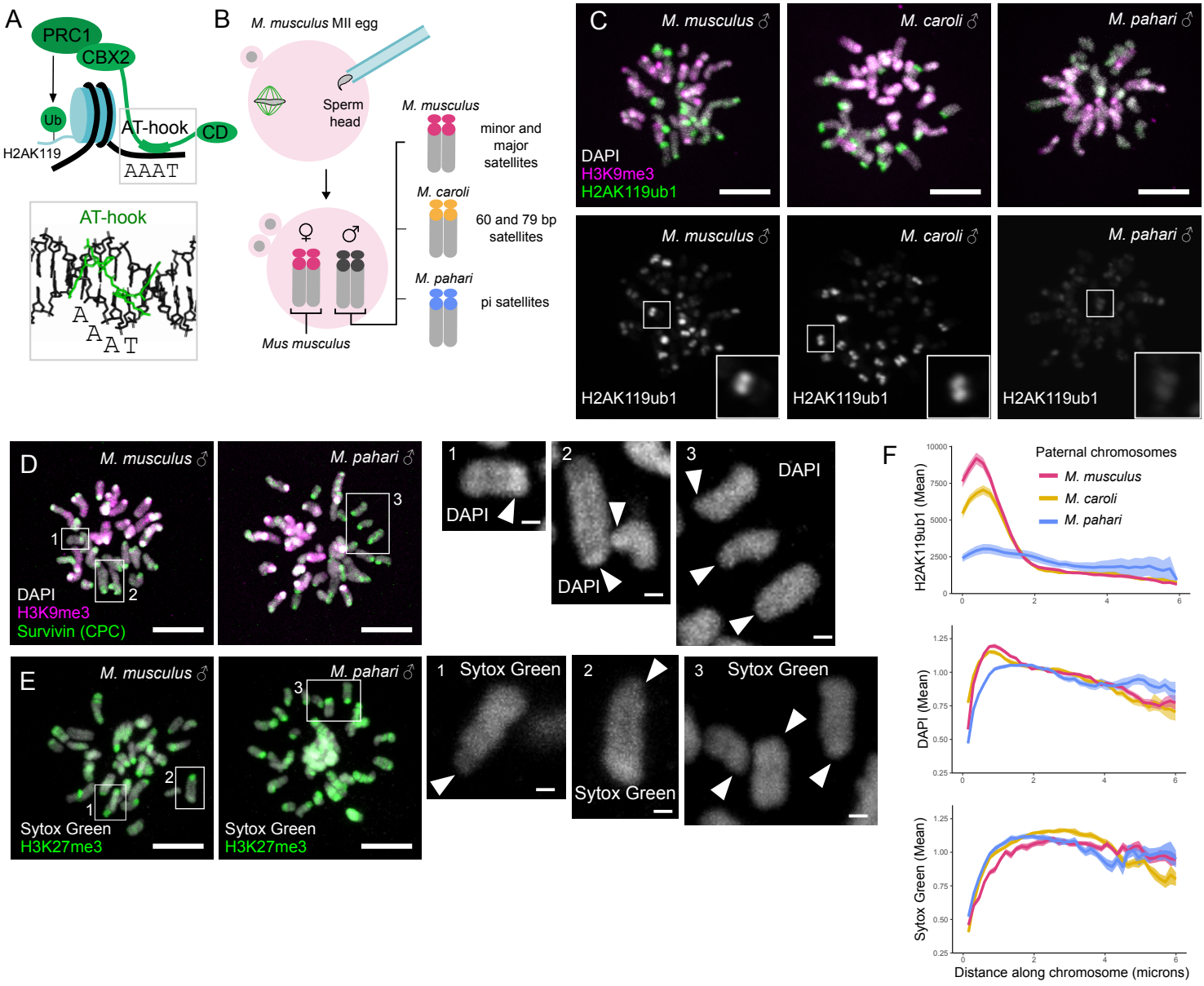

Figure 2

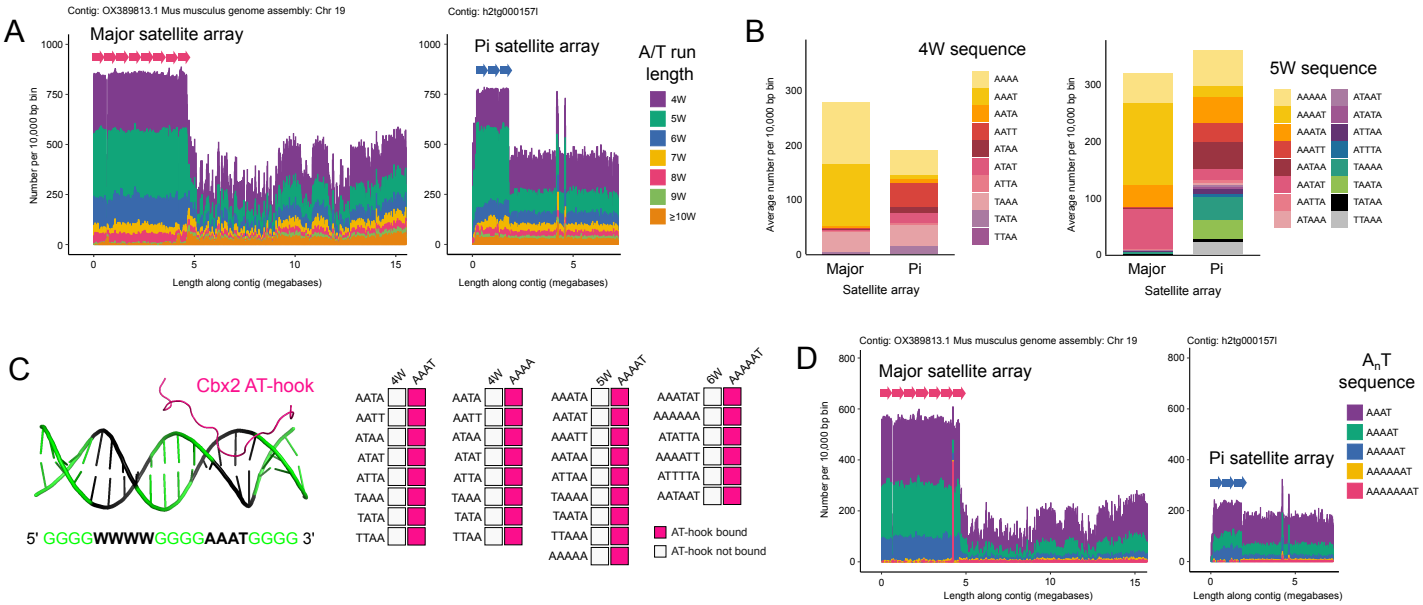

Figure 3