Satellite DNA sequence dictates pericentromere heterochromatin formation and function

Piero Lamelza[1], Malena Parrado[1], Kathleen Leara[2], Ahmed Z. Balboula[2], Thomas M. Keane[3], Michael A. Lampson[1,4*]

[1]Department of Biology, University of Pennsylvania, Philadelphia, PA, USA
[2]Division of Animal Sciences, University of Missouri, Columbia, MO, USA
[3]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK
[4]Penn Center for Genome Integrity, University of Pennsylvania, Philadelphia, PA, USA

*Address correspondence to: 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. To test the functional consequences of this rapid sequence evolution, we establish hybrid mouse embryos as a model system to compare divergent satellite arrays from distinct species in a common cytoplasm. We show that variation in satellite sequence impacts heterochromatin formation, recruitment of the Chromosome Passenger Complex (CPC), and interactions with the mitotic spindle. Differences in satellite DNA sequence alter pericentromere packaging by Polycomb Repressive Complex 1 (PRC1), as satellite arrays that recruit PRC1 are enriched for specific A/T-sequences that the PRC1 AT-hook preferentially binds. Furthermore, PRC1 heterochromatin modifies pericentromere function by inhibiting recruitment of the CPC, increasing microtubule forces on kinetochores during mitosis. Our results provide a direct link between satellite DNA composition and mitotic chromosome behavior and highlight early embryogenesis as a critical point in development that is sensitive to satellite DNA evolution.

Teaser

Rapidly evolving satellite DNA impacts heterochromatin formation and recruitment of a key regulator of chromosome segregation.

Introduction

Satellite DNA is composed of tandemly repeated DNA sequences that make up large non-coding regions of eukaryotic genomes, e.g. 7% and 8% of human and mouse genomes, respectively (*1*, *2*). Satellites are among the most rapidly evolving genomic sequences, with their monomer size, nucleotide sequence, genomic distribution and abundance varying widely between species (*3–6*). Recent advances in long-read sequencing technology have made huge progress in precisely characterizing the genetic composition and diversity of repetitive satellite arrays (*1*, *7–11*). However, it is still largely unknown what functional consequences may arise from the natural diversity of these sequences.

Despite their rapid evolution, satellites often underlie pericentromeres, heterochromatic loci adjacent to centromeres with conserved roles in chromosome segregation during cell division (*12*). This paradoxical observation is partially resolved by evidence indicating that pericentric heterochromatin maintenance and function are largely independent of satellite DNA sequence. Instead, factors that package and transcriptionally silence satellites rely on the recognition of H3K9me3 marked nucleosomes that are canonically enriched at pericentromeres (*13*). Once established, H3K9me3 can propagate heterochromatin through successive cell cycles by recruiting histone methyltransferases that add new H3K9me3 (*13*). Additionally, phosphorylation of pericentromeric histones during mitosis by Haspin and Bub1 kinases facilitates the recruitment of the Chromosome Passenger Complex (CPC) (*14*), a key regulator of kinetochore-microtubule attachments required for accurate chromosome segregation (*15*, *16*). Because these epigenetic mechanisms do not directly depend on DNA sequence, potential functional impacts of variation in satellite DNA composition have been difficult to determine (*17*).

Mouse embryos provide an attractive model system to test the significance of satellite DNA evolution for pericentromere heterochromatin formation and function. Because most histones are replaced by protamines in mouse sperm, the paternal genome must undergo a drastic repackaging into nucleosome-based chromatin after fertilization (*18*, *19*). These nucleosomes lack H3K9me3, creating a unique developmental window in which canonical epigenetic pathways of heterochromatin maintenance are absent on paternal satellites (*20*). Instead, Polycomb Repressive Complexes 1 and 2 (PRC1 and PRC2) package paternal satellites in heterochromatin marked with H2AK119ub1 and H3K27me3, respectively, which typically suppress gene expression elsewhere in the genome (*20–22*). Both modifications contribute to the transcriptional silencing of the 240 bp A/T-rich "major satellite" repeats at *M. musculus* pericentromeres (*20*, *23*), with PRC1 also crucial for maintaining the integrity of these arrays (*24*). Furthermore, both complexes can localize to pericentromeres by recognition of satellite DNA (*21*, *22*). The CBX2 subunit of PRC1

69 recognizes satellites via its AT-hook, a five amino acid motif with nanomolar binding affinity for
70 the relatively narrow minor grooves of double stranded DNA (dsDNA) generated by stretches of
71 tandem A/T nucleotides (A/T runs) (*25–29*) (Fig 1A). PRC2 localization to paternal
72 pericentromeres likely requires BEND3, a methylation sensitive DNA-binding protein that can
73 bind unmethylated major satellite DNA *in vitro* (*21*).
74 We developed a system to test whether variation in satellite DNA sequence alters paternal
75 heterochromatin by fertilizing *M. musculus* eggs with sperm from closely related mouse species
76 with divergent satellite repeats. We inject sperm into egg cells (intracytoplasmic sperm injection,
77 ICSI (*30*)) to bypass fertilization barriers between species, allowing us to test satellite DNA
78 variation that would otherwise be inaccessible by natural mating. We focus our analysis on the
79 one-cell zygote, before the major transcriptional activation of the mouse embryo (*31*), so that our
80 results reflect interactions between satellite sequences and maternally inherited *M. musculus*
81 Polycomb complexes rather than a mixture of subunits from both species. Analyzing
82 heterochromatin in the zygote also avoids the complication of potential genetic incompatibilities
83 that typically arise after the transcriptional activation of the hybrid genome (*32–34*).
84 We demonstrate that variation in satellite DNA sequence significantly influences PRC1
85 recruitment and heterochromatin formation at pericentromeres. Satellite arrays that robustly
86 recruit PRC1 are enriched for specific A/T-runs (i.e. AAAAT) that optimally bind the CBX2 AT-
87 hook based on previous reports of AT-hook proteins (*29*) and our AlphaFold3 modeling. We find
88 that PRC1 heterochromatin modifies pericentromere function by inhibiting the recruitment of the
89 Chromosome Passenger Complex (CPC) to pericentromeres. The CPC ensures accurate
90 chromosome segregation by regulating chromosome attachment to the mitotic spindle by
91 destabilizing kinetochore-microtubule interactions. Pericentromeres that form PRC1
92 heterochromatin exhibit longer sister-kinetochore distances during metaphase, consistent with a
93 reduction in CPC stabilizing kinetochore-microtubule interactions and increasing the force pulling
94 sister-kinetochores apart. Furthermore, we find that paternal chromosomes with PRC1
95 heterochromatin are more prone to missegregate during anaphase compared to maternal
96 chromosomes, which lack PRC1. Our results show that diverse pericentric DNA sequences lead
97 to functionally diverse pericentromeres through differential binding of heterochromatin forming
98 complexes.
99
100 **Results**
101 ***Variation in PRC1 localization and heterochromatin formation on divergent satellites***

102 Given that major satellite DNA can direct PRC1 and PRC2 heterochromatin formation, we
103 hypothesized that divergent pericentric satellite sequences might alter CBX2 or BEND3 binding
104 and by extension H2AK119ub1 or H3K27me3 formation, respectively. We fertilized *M. musculus*
105 eggs with sperm from mouse species harboring genetically divergent pericentric satellites (Fig
106 1B). We selected three species (*M. musculus, M. caroli* and *M. pahari*) that differ in their satellite
107 monomer sequence and not only in satellite copy number as seen in *M. musculus* subspecies (*2*,
108 *5*). We injected sperm heads into MII eggs, arrested the resulting zygotes at the first mitosis by
109 kinesin-5 inhibition, and measured H2AK119ub1 and H3K27me3 by immunofluorescence.
110 Because only maternal chromosomes have H3K9me3 at this stage, we used this mark to
111 distinguish maternal and paternal chromosomes (*20*).
112 We found that all three species' pericentromeres robustly acquire H3K27me3 (Fig S1A-
113 B). In contrast, H2AK119ub1 is enriched on paternal *M. musculus* and *M. caroli* pericentromeres
114 relative to chromosome arms, but not on *M. pahari* pericentromeres (Fig 1C and 1E). To test
115 whether these differences result from differences in PRC1 localization, we measured PRC1's
116 core-catalytic subunit, RNF2, by immunofluorescence. Consistent with the H2AK119ub1
117 enrichment, paternal *M. musculus* pericentromeres are strongly enriched for RNF2 relative to
118 chromosome arms whereas *M. pahari* chromosomes only exhibit a slight enrichment near the
119 very ends of their chromosomes, presumably near centromeres (Fig 1D and 1E).
120 In certain contexts, PRC1 is recruited to chromatin by the chromodomain of its CBX
121 subunit, which directly binds H3K27me3 (*35–37*), and CBX2 also contains a chromodomain (Fig
122 1A). However, the presence of H3K27me3 at *M. pahari* paternal pericentromeres (Fig S1A-B),
123 which lack H2AK119ub1 and RNF2 (Fig 1C-E), indicates that H3K27me3 is not sufficient for
124 PRC1 recruitment during the first zygotic mitosis. In contrast to mitosis, both H3K27me3 and
125 RNF2 are enriched at both *M. musculus* and *M. pahari* paternal pericentromeres during G2 (Fig
126 S1C-E), consistent with previous reports that the CBX2 chromodomain contributes to PRC1's
127 pericentric localization during zygotic interphase (*22*). The difference in RNF2 localization
128 between G2 and mitosis indicates that the CBX2 chromodomain does not effectively localize
129 PRC1 during mitosis. We speculate that Aurora B-mediated phosphorylation of H3S28 during
130 mitosis (*38*) prevents the CBX2 chromodomain from binding H3K27me3, similar to how Aurora
131 B-mediated phosphorylation of H3S10 inhibits HP1 chromodomain binding to H3K9me3 (*39*,
132 *40*). Therefore, the observed differences between *M. musculus* and *M. pahari* paternal
133 pericentromeres during mitosis, in both PRC1 recruitment and heterochromatin formation, likely
134 reflect variation among the divergent satellite DNA sequences in their ability to bind the CBX2
135 AT-hook.

136

137 ***Pericentromeres enriched for narrow DNA minor grooves form PRC1 heterochromatin***

138 To explain the differences in pericentric RNF2 and H2AK119ub1, we hypothesized that
139 *M. musculus* and *M. caroli* pericentromeric satellites, but not *M. pahari* satellites, are enriched for
140 narrow DNA minor grooves that bind the AT-hook of CBX2 to recruit the PRC1 complex.
141 Following a previously described approach (*2*), we used DAPI staining to compare the amount of
142 narrow minor grooves generated by A/T-runs at each species' pericentromere, as DAPI
143 specifically binds to these regions on the dsDNA molecule (*41–43*). Consistent with previous
144 observations (*2*, *9*, *44*), we find increased DAPI staining at *M. musculus* and *M. caroli*
145 pericentromeres relative to chromosome arms, but not at *M. pahari* pericentromeres (Fig 2A-B
146 and Fig S2A). To exclude the possibility that DAPI enrichment at pericentromeres represents a
147 relatively more compact chromatin state compared to the rest of the chromosome, we stained with
148 Sytox Green, an intercalating DNA-dye with no considerable DNA sequence preference (*45*).
149 None of the three species show enrichment of Sytox Green at their pericentromeres (Fig 2C-D
150 and Fig S2B). Together, these results demonstrate that *M. musculus* and *M. caroli*
151 pericentromeres are enriched for narrow A/T-rich minor grooves that can robustly bind the AT-
152 hook of CBX2, whereas *M. pahari* pericentromeres are not.

153

154 ***Satellite sequence determinants of PRC1 binding***

155 We considered two possibilities for the underlying differences in pericentric DNA
156 sequence between species, leading to differences in DAPI enrichment and PRC1 localization. We
157 initially focused on comparisons between *M. musculus* and *M. pahari* because of the long-read
158 sequencing data available for these species. First, *M. musculus* major satellites may have a higher
159 frequency of A/T runs compared to *M. pahari* pericentric satellites (pi satellites) (*9*). However,
160 major and pi satellite consensus sequences have similar frequencies of A/T runs of lengths
161 ranging from four (4W, the minimal number of nucleotides the AT-hook can span) to ten or more
162 (≥10W) (Fig S2C). To capture nucleotide diversity that may be lost in consensus sequences, we
163 also calculated the frequencies of A/T runs across genomic scaffolds containing either major or pi
164 satellite arrays. Consistent with consensus sequences, we find that both species' satellite arrays
165 have similar frequencies of A/T runs (~825 vs ~750 per 10 kb bin in major satellite or pi satellite,
166 respectively) (Fig 3A). The frequencies are also similar for each length of A/T runs (Fig 3A and
167 Fig S2D), with the exceptions of 6W and ≥10W, which are relatively more abundant in major and
168 pi satellite arrays, respectively. However, it is unclear how differences in these lengths
169 specifically would lead to higher PRC1 and DAPI at *M. musculus* major satellite arrays.

170 Therefore, it is unlikely that differences in A/T-run frequency explain the difference in PRC1
171 heterochromatin formation.
172 The second possibility is that *M. musculus* major satellite arrays might be enriched for
173 specific A/T sequences optimal for CBX2 AT-hook binding, compared to *M. pahari* pi satellite
174 arrays. We analyzed the sequence composition of 4W, 5W and 6W A/T runs, which make up
175 ~75% of the A/T stretches in both species' satellite arrays and therefore most potential AT-hook
176 binding sites. The vast majority of 4W sequences in *M. musculus* major satellite arrays are AAAT
177 (114 per 10 kb), AAAA (113 per 10 kb) or TAAA (39 per 10 kb) (Fig 3B and Fig S2E). *M.*
178 *pahari* pi satellite arrays have a similar frequency of TAAA sites (37 per 10 kb) but much lower
179 frequencies of AAAT and AAAA sites (7.2 and 46 per 10 kb, respectively). One of these
180 sequences, AAAT, generates one of the narrowest minor grooves compared to other
181 tetranucleotides (*46*), suggesting preferential binding to an AT-hook. In addition, the core motif
182 of the AT-hook (PRGRP) preferentially binds the minor groove of the sequence 5' AA(A/T)T 3',
183 reflecting optimal Van der Waals packing when adenine bases are on opposite sides of the A/T
184 run (*29*). Because of these structural features and their relative abundance in major satellite, we
185 hypothesized that the CBX2 AT-hook may preferentially bind the narrow minor groove of
186 AAAT.
187 To test the binding preferences of the CBX2 AT-hook, we used AlphaFold3 (*47*) to
188 perform *in silico* pairwise competitive binding assays for AAAT vs all other possible A/T
189 tetranucleotides (Fig 3C). We first confirmed that AlphaFold3 accurately models AT-hook
190 binding to dsDNA using the well characterized HMGA1 AT-hook bound to a fragment of the
191 IFN-ß promoter (*28*) (PDB ID 2EZD) as a reference. The predicted structure closely matched the
192 known structure (RMSD=0.48 Å), and in both structures, the AT-hook occupied the minor groove
193 of the A/T run (AAATT) (Fig S2F). Furthermore, the confidence scores of the predicted structure
194 (iPTM=0.48 and PTM=0.47) are similar to the models generated in our *in silico* competitive
195 binding assay (Table S1). The AT-hooks of CBX2 and HMGA1 share the same core sequence
196 (PRGRP) which binds the minor groove, supporting our use of AlphaFold3 to model the AT-hook
197 binding preference of CBX2.
198 For each competition assay, we model a single dsDNA molecule encoding two distinct
199 A/T runs flanked by four G:C base pairs (5' GGGGWWWWGGGGWWWWGGGG 3'). These
200 two A/T runs "compete" to bind a single peptide consisting of the core CBX2 AT-hook along
201 with flanking amino acids (RKRGKR**PRGRP**RKHTVTSS). All AlphaFold models predict that
202 the CBX2 AT-hook prefers to bind AAAT over almost all other A/T tetranucleotides (Fig 3C).
203 When AAAT and AAAA compete against each other, their relative order on the dsDNA dictates

204 which sequence binds the AT-hook (Fig S2G, see Materials and Methods), indicating that these
205 two sequences have similar affinity for the AT-hook. We obtain similar results when analyzing
206 5W and 6W A/T runs. AAAAT and AAAAAT are the most abundant 5W and 6W sequences in
207 the *M. musculus* major satellite array (147 per 10 kb and 70 per 10 kb, respectively) (Fig 3B and
208 Fig S2E). These sequences outcompete all other prevalent 5W and 6W sequences in both major
209 and pi satellite arrays for CBX2 AT-hook binding (Fig 3C and Fig S2G). Furthermore, AAAAT
210 and AAAAAT are 8 and 2.5-fold less frequent, respectively, in *M. pahari* pi satellite compared to
211 *M. musculus* major satellite arrays (Fig 3B and Fig S2E). We also find that AAAAT has one of
212 the narrowest minor grooves of the analyzed A/T-runs (Fig S3), consistent with this feature
213 enhancing AT-hook binding (*29*), indicating that the AlphaFold3 models capture empirical
214 principles governing AT-hook-DNA interactions.

215 Our AlphaFold3 modeling shows that the CBX2 AT-hook prefers to bind stretches of
216 adenine nucleotides that end with a single thymine (A_nT). To extend this analysis to include A_nT
217 sequences (like AAAT) present in longer A/T runs, we quantified the frequency of A_nT sequences
218 of various lengths (i.e. A_3T, A_4T, A_5T, A_6T and A_7T) in major and pi satellite arrays independent
219 of total A/T run length. We find that major satellite arrays have approximately three times more
220 A_nT sequences than *M. pahari* pi satellite arrays (Fig 3D and Fig S2H). Furthermore, pi satellite
221 arrays exhibit A_nT frequencies similar to chromosome arms. Within major satellite arrays, longer
222 A_nT sequences (i.e., A_4T and A_5T) are more enriched relative to chromosome arms (roughly 5.3-
223 fold and 6.8-fold, respectively) than A_3T (roughly 2.5-fold) (Fig 3D, Fig S2H and Table S2). Our
224 *in silico* analysis shows that A_4T and A_5T outcompete A_3T for CBX2 AT-hook binding (Fig 3C),
225 consistent with previous findings that longer A/T runs are preferred binding sites for AT-hook
226 proteins (*48*). Overall, these analyses indicate that PRC1 and its H2AK119ub1 modification are
227 enriched at *M. musculus* major satellite arrays because they are enriched for longer A_nT sequences
228 that preferentially bind the CBX2 AT-hook. In contrast, *M. pahari* pi satellite arrays have fewer
229 and shorter A_nT sequences, like chromosome arms, and fail to enrich both PRC1 and
230 H2AK119ub1.

231 To further test our model, we compared PRC1 recruitment between *M. musculus* and *M.*
232 *spretus* paternal pericentromeres. Both species harbor minor and major satellite arrays, but in
233 roughly inverse proportions. Major and minor satellites make up roughly ~7.9% and ~0.5% of the
234 *M. musculus* genome, respectively, compared to ~0.2% and ~4.4% in *M. spretus* (Fig 3E) (*2*).
235 Although long-read assemblies of *M. spretus* are not currently available, the minor satellite
236 consensus sequence is not enriched for A_nT sequences (Fig 3E). We therefore predicted that the
237 abundant minor satellites at *M. spretus* pericentromeres would not recruit PRC1 as strongly as the

major satellites enriched *M. musculus* pericentromeres. Indeed, we find that paternal *M. musculus* pericentromeres recruit roughly 2-fold more RNF2 than paternal *M. spretus* pericentromeres (Fig 3F-G). This result demonstrates that the A_nT-enriched major satellite is the primary driver of PRC1 recruitment to *M. musculus* pericentromeres, based on direct comparison of species whose pericentromeres primarily differ in the relative abundance of the same two satellite sequences.

AlphaFold modeling predicts a novel A_nT-rich Mus caroli satellite

Because *M. caroli* pericentromeres form PRC1 heterochromatin (Fig 1C), our model predicts *Mus caroli* satellites should be enriched for A_nT sequences. However, analysis of the known *M. caroli* satellites (*49*) shows that the 60-bp satellite lacks A_nT sequences, while the 79-bp satellite contains only A_3T at a low frequency comparable to *M. pahari* pi satellite (~200 per 10 kb) (Fig S4A). Thus, these *M. caroli* satellites are not enriched for the longer A_nT sequences found in major satellite. We therefore hypothesized that *M. caroli* has additional, previously uncharacterized satellites containing longer A_nT sequences that are responsible for PRC1 recruitment and heterochromatin formation.

Using an established computational pipeline (TAREAN) (*50*) with published *M. caroli* short-read sequences (see Materials and Methods), we identified two read clusters with a high probability of being satellites. The first cluster makes up 5.3% of input reads, with a consensus sequence identical to the known 79-bp satellite (Fig 4A) (*49*). Reads corresponding to both the 60-bp and 79-bp known satellites are grouped within this cluster, likely reflecting their high sequence similarity (*49*) (Fig S4A, B, and E and Table S3). The second cluster makes up 6% of input reads, with a 59-bp consensus sequence (Fig 4A) that includes two A_4T runs and is more similar to *M. musculus* major satellite sub-repeats than to the *M. caroli* 60-bp and 79-bp satellites (Fig S4D). Arrays of the 59-bp consensus satellite would have a frequency of $A_{\geq 4}T$ sequences similar to major satellite arrays (i.e., 338 vs ~280 per 10 kb) (Fig 4A and Table S2). In contrast to the 60-bp/79-bp cluster, however, the 59-bp cluster includes many variants, suggesting a relatively diverse satellite family (Fig S4C and Table S4). To capture this sequence diversity in our analysis, we quantified $A_{\geq 4}T$ in reads mapped to a concatenated set of 59-bp satellite variants outputted by TAREAN (see Materials and Methods). The observed $A_{\geq 4}T$ frequency (0.030) closely matches the predicted value from the consensus (0.034) (Fig 4B) and is 4.3-fold higher than a random sampling of reads (0.007), similar to the enrichment of $A_{\geq 4}T$ sequences in major satellite compared to chromosome arms (Fig 3D and Table S2). Furthermore, reads mapped to a concatenated 60-bp/79-bp sequence show a very low $A_{\geq 4}T$ frequency (0.0008), consistent with the consensus sequence of both satellites not encoding any $A_{\geq 4}Ts$. Together these results reveal a 59-

272 bp satellite family in *M. caroli* with a frequency of longer A_nT sequences similar to major
273 satellite.

274 The 79-bp and 60-bp *M. caroli* satellites possess hallmarks of centromeric satellites, such
275 as high sequence homogeneity (*1*, *51*) based on k-mer analysis (Fig S4B and Table S3) and
276 CENP-B box motifs in the 79-bp satellite (*49*). We therefore hypothesized that these two satellites
277 underly centromeres, whereas the relatively diverse 59-bp satellites primarily make up
278 pericentromeres. To visualize 59-bp satellite arrays on *M. caroli* chromosomes, we injected *M.*
279 *musculus* / *M. caroli* hybrid zygotes with mRNA encoding dCas9:mCherry and a cocktail of
280 gRNAs designed against reads within the 59-bp satellite cluster (Fig S4E) (see Materials and
281 Methods). To ensure the specificity of our gRNAs, PAM sites are present in 59-bp satellites but
282 absent in both the 60-bp and 79-bp satellites (Fig S4E). We find that dCas9:mCherry localizes to
283 paternal *M. caroli* metaphase chromosomes but not to maternal *M. musculus* pericentromeres,
284 despite the sequence similarity between the 59-bp satellite and *M. musculus* major satellite (Fig
285 4C), showing the specificity of our gRNAs. Moreover, mCherry signals are located near
286 centromeric ends of *M. caroli* chromosomes, consistent with our prediction that the 59-bp satellite
287 is pericentromeric.

288 As an independent approach to localize 59-bp satellite arrays on *M. caroli* chromosomes,
289 we generated a *de novo M. caroli* genome assembly from long PacBio HiFi and Hi-C reads. Our
290 assembly includes 20 chromosome-sized scaffolds, each aligning well to one of the 20
291 chromosomes of a previously published short-read *M. caroli* assembly (CAROLI_EiJ_v1.1) (Fig
292 S5A), indicating we have assembled all chromosome arms. Our assembly contains complete
293 copies of 99.6% of 9226 universal mammalian single copy protein coding genes (Fig S5B),
294 indicating that it is relatively complete in terms of gene content. Almost all chromosome-sized
295 scaffolds begin with a short 59-bp satellite array, and all end with a short telomere array (Fig 4D
296 and Fig S5C). Their positions at the ends of chromosome-sized scaffolds show that 59-bp
297 satellites are likely pericentromeric. 59-bp satellites can be interrupted by other sequences,
298 resulting in multiple discrete arrays that range in size from 400-bp to 15-kb (Fig S5D). Because
299 these small arrays fall within the average length of a PacBio read (~15 kb), it suggests they are
300 assembled by a few long reads anchoring them to unique sequence in chromosome arms. In
301 contrast, the majority of 59-bp satellite DNA (predicted by TAREAN to comprise 6% of the *M.*
302 *caroli* genome, Fig 4A) is likely largely unassembled. Consistent with this inference, 60-bp and
303 79-bp satellite arrays are not present in these chromosome-sized scaffolds and likely reside in
304 unassembled centromeres.

305 We quantified A_nT frequencies in the assembled 59-bp arrays and found that they are
306 similar to predictions from the consensus sequence (Fig 4E). We directly compared the average
307 number of A_nT sequences per 10 kb between the different satellite arrays and find that both major
308 satellite and 59-bp arrays have a higher enrichment of $A_{\geq 4}T$ sequences compared to chromosome
309 arms (4.6-fold and 4.3-fold, respectively) than pi satellite arrays (1.6-fold) (Fig 4F). Overall, our
310 analyses of four species' satellite sequences combined with their differing abilities to recruit
311 PRC1 indicate that longer A_nT sequences drive formation of PRC1 heterochromatin at paternal
312 pericentromeres in the first mitosis.
313
314 ***Consequences of PRC1 heterochromatin for pericentromere function***
315 Next, we asked if the difference in PRC1-heterochromatin between paternal *M. musculus*
316 and *M. pahari* chromosomes has consequences for pericentromere function. Notably,
317 H2AK119ub1 is positioned close to the site of another post-translational modification:
318 phosphorylation of H2AT121 by Bub1 kinase. H2AT121phos recruits the Chromosome
319 Passenger Complex (CPC) to pericentromeres via Sgo1 (Fig 5A) to ensure accurate chromosome
320 segregation by regulating kinetochore-microtubule interactions during mitosis (*14–16*). ChIP-seq
321 and *in vitro* assays on reconstituted nucleosomes have shown that H2AK119ub1 and
322 H2AT121phos are mutually exclusive (*52*), suggesting that one modification physically blocks
323 the deposition of the other on the same H2A C-terminal tail. We hypothesized that H2AK119ub1
324 occludes Bub1 kinase from phosphorylating H2AT121 and thereby reduces CPC binding. This
325 hypothesis predicts low H2AT121phos and low CPC on paternal *M. musculus* pericentromeres
326 (Fig 5A), which have high H2AK119ub1 compared to maternal *M. musculus* and paternal *M.*
327 *pahari* pericentromeres (Fig 1C and 1E).
328 To test this prediction, we first visualized H2AT121phos, the CPC subunit Survivin, and
329 the PRC1 subunit RNF2. In *M. musculus*♀/*M. musculus*♂ zygotes arrested at metaphase by
330 chemical inhibition of the anaphase promoting complex/cyclosome (APC/C), we found that
331 H2AT121phos is present along the entire chromosome (Fig 5B) as reported in the previous cell
332 cycle in mouse oocytes (*53*, *54*). Moreover, we found an inverse relationship between RNF2 and
333 H2AT121phos, with the presence of RNF2 at paternal pericentromeres coinciding with a
334 reduction in H2AT121phos (Fig 5B). In contrast, maternal pericentromeres with undetectable
335 RNF2 show no reduction in H2AT121phos. Consistently, when we compared regions of maternal
336 and paternal chromosomes that both lack RNF2, H2AT121phos intensities were similar,
337 indicating that reduced H2AT121phos depends on RNF2. On average, H2AT121phos is reduced
338 to ~60% at paternal *M. musculus* pericentromeres compared to maternal pericentromeres (Fig

339 5B). Consistent with H2AK119ub1 heterochromatin inhibiting H2AT121phos, Survivin intensity
340 on paternal *M. musculus* pericentromeres is reduced to ~65% compared to paternal *M. pahari* and
341 maternal *M. musculus* pericentromeres, both of which form relatively less H2AK119ub1 (Fig 5C).
342 Survivin is also reduced on paternal *M. caroli* pericentromeres, which have H2AK119ub1,
343 compared to maternal *M. musculus* pericentromeres (Fig S6). Together, these results show that
344 PRC1 heterochromatin on paternal pericentromeres inhibits the acquisition of H2AT121phos and
345 thereby reduces CPC recruitment.

346 The Aurora B kinase subunit of the CPC phosphorylates kinetochore proteins to
347 destabilize kinetochore-microtubule interactions (*55*). In tissue culture cells, expression of a non-
348 phosphorylatable mutant of a key Aurora B substrate, Hec1, results in increased stabilization of
349 kinetochore-microtubule interactions and increased forces pulling the sister kinetochores apart at
350 metaphase (*56*). Increased inter-kinetochore distance therefore provides a functional readout for
351 reduced Aurora B activity (Fig 5A). To test for functional consequences of CPC differences
352 between pericentromeres in the zygote, we measured distances between sister kinetochores at
353 metaphase. We found increased inter-kinetochore distances for paternal *M. musculus*
354 chromosomes compared to either maternal *M. musculus* or paternal *M. pahari* chromosomes (Fig
355 5D). This result is consistent with our hypothesis that PRC1 heterochromatin inhibits H2AT121
356 phosphorylation, downregulates the CPC, and stabilizes kinetochore-microtubule interactions on
357 paternal *M. musculus* chromosomes. Increased inter-kinetochore distances might also reflect
358 reduced pericentric cohesion because in addition to recruiting the CPC, Sgo1 also recruits
359 phosphatases that counteract cohesin removal that occurs along chromosome arms during mitotic
360 prophase (*57*, *58*) (Fig 5A). However, sister chromatids appear in close proximity along most of
361 their lengths during zygotic mitosis (*59*) (Fig 5D), suggesting that cohesins are not removed
362 during prophase at this stage of development. Therefore, we interpret the increase in inter-
363 kinetochore distance as a consequence of reduced CPC activity.

364 The known function of the CPC in regulating kinetochore-microtubule interactions
365 suggests that the observed differences between maternal and paternal pericentromeres may lead to
366 differences in the frequency of chromosome segregation errors in anaphase (e.g., due to a
367 kinetochore connected to microtubules emanating from both spindle poles) (*60*). To test this
368 possibility, we performed live imaging of pure-species *M. musculus* zygotes, in which paternal
369 pericentromeres form PRC1 heterochromatin and recruit less CPC than maternal pericentromeres.
370 The paternal pronucleus is larger than the maternal pronucleus in late G2, allowing us to
371 distinguish and automatically track paternal and maternal chromosomes during the first mitosis
372 (Movies S1 and S2). We observed at least one lagging chromosome in ~25% of anaphases

373 (41/162) (Fig 5E). In most cases (37 out of 41), we were able to determine the parental origin and
374 found that paternal chromosomes are over-represented (24/37=65%) compared to the null
375 prediction of 50% (*P=0.049*). This result indicates that paternal chromosomes are more prone to
376 segregation errors in *M. musculus* zygotes, consistent with PRC1-enriched pericentric
377 heterochromatin compromising chromosome segregation fidelity.
378

379 **Discussion**

380 Long-read sequencing technology continues to reveal the extensive diversity of satellite
381 DNA sequence composition (*8*, *9*, *11*, *61*), but experimental systems that test the functional
382 impact of this diversity remain limited. In the mouse model species *M. musculus,* centromeric and
383 pericentromeric satellite variation is limited to copy-number differences of the same satellite
384 sequences (*2*, *5*, *53*). Our hybrid embryo system overcomes this limitation by incorporating
385 sequence variation between satellites that is only found in more divergent non-model mouse
386 species.

387 Using this system, we demonstrated that *M. musculus* major satellite sequence is more
388 effective than *M. pahari* pi satellite sequence at recruiting PRC1 during mitosis. Consistent with
389 the known sequence preferences of AT-hook proteins (*29*), AlphaFold3 modeling indicates that
390 this difference results from the AT-hook of CBX2^{musculus} preferentially binding the exceptionally
391 narrow minor groove of long A_nT sequences, which are enriched in *M. musculus* major satellite
392 arrays when compared to *M. pahari* pi satellite arrays and chromosome arms. Furthermore, our
393 satellite DNA based model of PRC1 recruitment accurately predicted a novel A_nT-rich pericentric
394 satellite (59-bp satellite) in *Mus caroli* as well as reduced PRC1 recruitment to A_nT-depleted
395 minor satellites in *M. spretus.* Together, these results indicate that the enrichment of narrow minor
396 grooves generated by long A_nT sequences largely dictates PRC1 recruitment during mitosis,
397 although other species-specific sperm features (e.g. chromatin packaging, histone variants, RNAs)
398 could also contribute.

399 Our finding of divergence between *Mus* species in the ability of pericentric satellites to
400 recruit PRC1 during mitosis is surprising, given the importance of PRC1 in maintaining *M.*
401 *musculus* paternal pericentromere integrity during the first zygotic anaphase (*24*). We propose
402 that PRC1 localization to *M. pahari* pericentromeres during interphase, rather than its persistence
403 on mitotic chromosomes, is sufficient for pi satellite array stability during mitosis. Consistent
404 with this possibility, CBX2 recruits DAXX and promotes H3.3 deposition at paternal
405 pericentromeres during zygotic interphase (*24*). PRC1 silencing of satellite transcription also
406 occurs during interphase (*20*). This model is consistent with our observation that all tested species

407 establish H3K27me3 at paternal pericentromeres, suggesting that PRC1 enrichment during
408 zygotic interphase through the chromodomain of CBX2 is conserved in *Mus.* Alternatively, *M.*
409 *musculus* major satellites and *M. caroli* 59-bp satellites may have specific properties that require
410 their packaging in PRC1 heterochromatin during mitosis, whereas *M. pahari* pi satellites can rely
411 on PRC2-mediated H3K27me3.

412 As another alternative model, the *M. pahari* egg cytoplasm may have specific adaptations
413 that allow PRC1 to be recruited to mitotic pericentromeres by a different mechanism. Because the
414 core CBX2 AT-hook sequence is conserved across Muridae (Fig S7), it is unlikely that the AT-
415 hook of CBX2^{pahari} has adapted to bind A/T sequences enriched at *M. pahari* pi satellites.
416 Similarly, the CBX2 chromodomain is identical across Muridae (data not shown), indicating that
417 the chromodomain of CBX2^{pahari} has not adapted to have a higher binding affinity towards
418 H3K27me3. However, vertebrates encode multiple PRC1 complexes that share the core ubiquitin
419 ligase subunits (RING1A/1B) but differ in associated proteins that diversify PRC1 function,
420 including how the complex is recruited to chromatin (*62*). Thus, another chromatin targeting
421 subunit expressed in *M. pahari* zygotes might recognize paternal *M. pahari* pericentromeres
422 during mitosis.

423 PRC1 localization to *M. pahari* paternal pericentromeres during G2 but not mitosis
424 suggests that binding of the CBX2 chromodomain to H3K27me3 is cell-cycle dependent. We
425 speculate that Aurora B-mediated phosphorylation of H3S28 during mitosis (*38*) prevents the
426 CBX2 chromodomain from binding H3K27me3, leading to loss of PRC1 from *M. pahari*
427 pericentromeres, which lack A_nT-rich satellites. Cytosolic deubiquitinating enzymes would then
428 remove H2AK119ub1 after nuclear envelope breakdown (*63*). This removal, together with our
429 finding that H2AK119ub1 inhibits CPC recruitment, provides a mechanism for pericentric
430 satellites to impact paternal chromosome segregation during the first zygotic mitosis. A_nT-rich
431 satellites at paternal *M. musculus* and *M. caroli* pericentromeres maintain PRC1 in mitosis
432 through the CBX2 AT-hook, leading to low CPC. In contrast, low PRC1 at paternal *M. pahari*
433 pericentromeres leads to high CPC, similar to maternal *M. musculus* pericentromeres, where
434 H3K9me3-based heterochromatin inhibits PRC1 recruitment (*20*, *22*).

435 Two lines of evidence support the functional significance of these differences in mitotic
436 PRC1. First, low CPC is expected to stabilize kinetochore microtubules, leading to increased
437 sister-kinetochore separation as we observe during the first zygotic metaphase for paternal *M.*
438 *musculus* chromosomes (high PRC1) compared to maternal *M. musculus* or paternal *M. pahari*
439 chromosomes (low PRC1). Looser chromatin packing due to CBX2/PRC1 retention on mitotic
440 chromosomes could also contribute to the increased sister-kinetochore separation. Second, we

441 find that paternal *M. musculus* chromosomes lag significantly more than maternal *M. musculus*
442 chromosomes during zygotic anaphase, indicating more frequent kinetochore-microtubule
443 attachments errors with high PRC1 and low CPC.
444 We propose that compromised paternal chromosome segregation fidelity, due to the
445 PRC1-mediated decrease in CPC recruitment (and/or decrease in chromatin compaction), selects
446 against large A_nT-rich satellite arrays in natural populations. Consistent with this idea, wild-
447 caught *M. musculus* have roughly ten times fewer major satellite repeats than inbred laboratory
448 strains, where natural selection is relatively weak (*5*). However, pericentric PRC1
449 heterochromatin is required for the transcriptional silencing and stability of paternal major
450 satellite arrays (*22*, *24*), suggesting an evolutionary trade-off between proper packaging of major
451 satellites and chromosome segregation fidelity.
452 To avoid this trade-off in most cell cycles, we propose that H3K9me3 heterochromatin
453 serves to prevent DNA sequence-based pathways of PRC1 recruitment, preserving pericentric
454 CPC during mitosis and chromosome segregation fidelity. This epigenetic heterochromatin
455 suppresses functional diversity that might otherwise arise from differences in satellite DNA
456 sequence composition. In contrast, our findings highlight the first zygotic cell cycle as a unique
457 environment where heterochromatin formation and function are sensitive to satellite DNA
458 sequence evolution due to the loss of H3K9me3 from the paternal genome during spermiogenesis.
459

460 **Materials and Methods**

461 ***Mouse strains***

462 The mouse strain representing *Mus musculus* in our ICSI experiments is FVB/NJ,
463 purchased from Jackson Laboratory (strain#001800). FVB/NJ was chosen because its MII eggs
464 can survive intracytoplasmic sperm injection and because it is an inbred strain. *M. caroli*
465 (CAROLI/EiJ, strain#000926) and *M. pahari* (PAHARI/EiJ, strain #002655) were also purchased
466 from Jackson Laboratory. We maintain our own colony of *M. pahari* because Jackson Laboratory
467 no longer carries this species. *M. spretus* was purchased from RIKEN BioResource Research
468 Center (SPR2, RBRC00208). The mouse strain representing *Mus musculus* in our live imaging
469 experiments, CF-1, was purchased from Envigo, Inotiv, Indianapolis, IN, USA (Hsd:NSA). All
470 animal experiments and protocols were approved by and conducted according to the regulations
471 of the University of Pennsylvania (Institutional Animal Use and Care Committee, protocol
472 #804882) or the University of Missouri (Animal Care Quality Assurance Ref. # 9695).
473

474 ***Intracytoplasmic sperm injection (ICSI)***

475 To prepare sperm for ICSI, the epididymis and vas deferens were removed from males and
476 dissected in PBS to release mature sperm. Sperm were allowed to swim-out for 10 mins on a
477 37°C slide warmer. After swim out, sperm were pelleted by centrifugation at 700xg at 4°C for 5
478 min. The PBS supernatant was removed, and sperm were then washed twice with ice-cold
479 Nuclear Isolation Media (NIM, 123 mM KCl, 2.6mM NaCl, 7.8 mM Na_2PO_4, 1.4 mM KH_2PO_4, 3
480 mM EDTA, pH adjusted to 7.2 using 1M KOH) with 1% poly vinyl alcohol (PVA). After the
481 second wash, sperm were resuspended in 100 μL of NIM 1% PVA and then sonicated using a
482 Branson Sonic Bath Model 1210 for 15-20 second intervals until at most 30% of sperm had their
483 heads detached from their tails. Sonicated sperm were washed twice with NIM 1% PVA. After
484 the second wash, the sperm pellet was resuspended in a 1:1 solution of glycerol and NIM 1%
485 PVA and placed at -20°C until the day of injection (the following day or 1 week later).
486 Females were super-ovulated by injection with 5 U PMSG (Peptides International)
487 followed by injection with 5 U hCG (Sigma) 48 hours later. MII eggs were collected from
488 females 14-15 hours post-hCG injection and placed in M2 media (Sigma) with hyaluronidase
489 (0.15 mg/mL) to remove cumulus cells. MII eggs were then washed through four drops of M2
490 media supplemented with 4 mg/mL BSA (M2+BSA) and left on a 37°C plate warmer until
491 injection. 10-20 μLs of sonicated sperm was diluted in 200 μL of NIM 1% PVA, gently vortexed,
492 and then washed twice with NIM 1% PVA. After the last wash, the sperm pellet was resuspended
493 in 100 μL of NIM 1% PVA and left on ice until injection.
494 Batches of 10 MII eggs were injected with sperm heads at a time in room temperature
495 drops of M2+BSA. After injection, each batch was moved to a drop of M2+BSA on a 37°C plate
496 warmer and allowed to recover for 1 hour. After recovery, each batch was washed through four
497 drops of pre-equilibrated AKSOM (Millipore Sigma) and placed in at 37°C humidified incubator
498 with 5% CO_2 until injections were complete. Depending on the experiment, at least three hours
499 after injection, eggs were then moved into a drop of AKSOM with either 10 μM STLC (kinesin-5
500 inhibitor, Sigma) (to arrest cells in mitosis with monopolar spindles) or 5 μM proTAME (APC/C
501 inhibitor, R&D Systems) (to arrest zygotes in metaphase), then allowed to develop overnight in a
502 37°C incubator with 5% CO_2. Zygotes arrested in mitosis were fixed in 2% paraformaldehyde 19
503 hours post-sperm injection. To stain for RNF2 and H3K27me3 in G2 zygotes, injected eggs were
504 moved to AKSOM and then allowed to develop 14 hours post-sperm injection, at which point
505 they were fixed in 2% paraformaldehyde.
506
507 ***Fixing and staining***

508 Zygotes were fixed in 37°C 2% paraformaldehyde in PBS for 20 minutes, washed through
509 three pools of blocking solution (PBS containing 0.5% BSA, 0.01% Tween-20) and then left at
510 4°C overnight. The next morning cells were permeabilized in PBS containing 0.5% Triton X-100
511 (Sigma) for 15 minutes at room temperature, quickly washed through two pools of blocking
512 solution and then allowed to block for 20 minutes at room temperature. For H3K9me3 and
513 H3K27me3 staining, the cells were treated with lambda-phosphatase (1600 U, NEB) for 1 hour at
514 37°C in a humidified chamber. Phosphatase treatment is required to detect H3K9me3 and
515 H3K27me3 staining on mitotic chromosomes, presumably by removing adjacent H3S10 and
516 H3S28 phosphorylation (*64*). Embryos were then quickly washed through two pools of blocking
517 solution and then incubated with primary antibodies for 1 hour in a dark humidified chamber at
518 room temperature (or overnight at 4°C for H2AT121phos staining). Afterwards, embryos
519 underwent three 15-minute washes in blocking solution and then were incubated with secondary
520 antibodies for 1 hour in a dark humidified chamber at room temperature. Cells then underwent
521 three 15-minute washes in blocking solution and were mounted in Vectashield with DAPI
522 (Vector) to stain chromosomes. For cells that were stained with Sytox Green (Invitrogen), the first
523 wash after incubation with secondary antibodies included 1 uM Sytox Green and after the next
524 two washes were mounted in Vectashield without DAPI. Primary antibodies used are rabbit anti-
525 H2AK119ub1 (1:800, Cell Signaling, D27C4), rabbit anti-H3K27me3 (1:700, Cell Signaling,
526 C36B11), mouse anti-H3K9me3 (1:200, Active Motif, 39285), rabbit anti-H3K9me3 (used when
527 co-staining against RNF2, 1:500, Active Motif, 39162), rabbit anti-Survivin (1:500, Cell
528 Signaling, 71G4B7), rabbit anti-H2AT120phos (1:2500, Active Motif, 39392), mouse anti-
529 RNF2/Ring1B (1:500, Active Motif, 39664) and rabbit anti-CENP-C(*65*) (1:500). Secondary
530 antibodies used are donkey anti-rabbit Alexa Fluor 488, donkey anti-mouse Alexa Fluor 594,
531 donkey anti-rabbit Alexa Fluor 594, and donkey anti-mouse Alex Fluor 647. All secondary
532 antibodies are purchased from Invitrogen and used at a dilution of 1:500.
533
534 ***Microscopy***
535 Some confocal images were collected as z-stacks with 0.5 µm intervals, using a
536 microscope (DMI4000 B; Leica) equipped with a 63× 1.3 NA glycerol-immersion objective lens,
537 an xy piezo Z stage (Applied Scientific Instrumentation), a spinning disk confocal scanner
538 (Yokogawa Corporation of America), an electron multiplier charge-coupled device camera
539 (ImageEM C9100–13; Hamamatsu Photonics), and either an LMM5 (Spectral Applied Research)
540 or Versalase (Vortran Laser Technology) laser merge module, controlled by MetaMorph software
541 (Molecular Devices, v7.10.3.294). Confocal images were also collected using a Leica TCS SP8

542 Four Channel Spectral Confocal System with a 63x objective lens. A z-stack of 0.3 µm intervals
543 was used to collect images of embryos used to calculate sister-kinetochore distances. All samples
544 in an experiment were imaged using the same laser settings.
545
546 ***Image quantification and statistical analysis***
547 All image analysis was carried out using ImageJ/Fiji (*66*). To measure signal intensities
548 along single chromosomes, we selected chromosomes that were mostly in the XY plane and
549 generated max intensities projections of them. Beginning from the pericentric end, we then used
550 the segmented line tool to manually draw a linear ROI (roughly the width of the chromosome)
551 along the chromosome's length and measured the average signal intensity per unit distance along
552 the ROI. Average background was then subtracted from these values. To more easily compare the
553 distributions of DAPI and Sytox Green intensity, their values were normalized to the average
554 intensity within each cell/embryo using custom R scripts.
555 To measure Survivin intensity at pericentromere ends, we also selected chromosomes that
556 were mostly in the XY plane and generated maximum intensity projections of them. We then use
557 the rectangular selection tool to generate square ROIs that encompassed pericentric ends of
558 chromosomes and measured average intensity within the ROI. Average background was then
559 subtracted from these values. The resulting values were then normalized to maternal
560 chromosomes in each cell/embryo using custom R scripts.
561 To measure the distance between sister-centromeres, we uniquely labelled each
562 centromere and extracted its 3D coordinates using 3D maxima finder and 3D Roi Manager (based
563 on CENP-C signal). We then manually assigned sister-centromeres with the help of
564 BigDataViewer to "rotate" the image stack in cases where the spindle pole axis was not in the XY
565 plane and then used custom R scripts to calculate the distance between them.
566 *P*-values were determined using R. We performed non-parametric Kruskal-Wallis tests
567 followed by Dunn's test and used the Bonferroni method to correct *P*-values for multiple testing.
568 Plots were made in the ggplot2 package in R.
569
570 ***Satellite sequence analysis***
571 *M. musculus* scaffolds containing major satellite arrays were found by BLASTing the
572 major satellite consensus sequence (*5*) on NCBI BLAST server. These scaffolds are from a long-
573 read *M. musculus* genome assembly generated by the Mouse Genome Project and Darwin Tree of
574 Life Project (GenBank: GCA_947593165.1) and were downloaded from the National Center for
575 Biotechnology (NCBI). The pi satellite array containing scaffolds are from a published *M. pahari*

576 long-read assembly (*9*). These scaffolds were chosen because they represent relatively complete
577 assemblies of functional *M. pahari* centromeres and pericentromeres based on CENP-A and
578 H3K9me3 ChIP-seq. We chose to study major and pi satellite arrays because they are the most
579 abundant pericentric satellites and because PRC1 is known to associate with major satellite (*20*).
580 To quantify the frequency of A/T runs of various lengths along a scaffold, we used the
581 matchPattern() function in the Biostrings package in R
582 (https://bioconductor.org/packages/Biostrings). Specifically, we searched for A/T runs that were
583 at least four nucleotides long (i.e. matched "WWWW"), with overlapping matches eventually
584 being merged into one longer A/T run by using the reduce() function. To analyze the composition
585 of 4W and 5W A/T runs, we used the matchPattern() function to find all possible 4W and 5W
586 sequences and then only kept matches that overlapped with 4W or 5W A/T run lengths,
587 respectively. To analyze the composition of 6W A/T runs, we used the 6W A/T run coordinates to
588 extract sequences within the satellite array using the Genomic Ranges package (*67*) in R.
589 Sequences that are reverse complements of each other are redundant and therefore were merged
590 into one. The average number of times a particular sequence occurred per 10 kb bin within an
591 array was calculated by dividing the total number of occurrences of a particular sequence within
592 the array by the length of the array and then multiplying by 10,000. We defined the range of the
593 satellite arrays for each scaffold based on coordinates from satellite consensus sequence BLAST
594 results and chromosome arms were defined as non-satellite regions of the same assembled
595 scaffolds. Finally A_nT sequences were also found along scaffolds using the matchPattern()
596 function. To prevent over-counting of smaller AnT sequences (e.g. overlap of AAAT and
597 AAAAT), only the longest of overlapping A_nT sequences were counted.
598
599 ***AlphaFold3 modeling***
600 All sequences were modeled using the AlphaFold3 webserver
601 (https://alphafoldserver.com/). To model a single dsDNA molecule encoding two distinct A/T
602 runs flanked by four G:C base pairs, we entered a single forward and a single reverse strand (e.g.
603 forward: 5'GGGGAAAAGGGGAAATGGGG 3', reverse: 5'CCCCATTTCCCCTTTTCCCC 3').
604 We also modeled a single peptide encoding the AT-hook of CBX2 with flanking amino acids
605 (RKRGKR**PRGRP**RKHTVTSS). Alphafold3 outputs five model structures (0-4), with 0 scored
606 as the best and the fourth scored as the worst. Each model was visually inspected using Pymol
607 (The Pymol Molecular Graphics System, Version 3.0 Schrodinger, LLC). All five models for
608 each competition assay were consistent with one another (Table S1). To rule out the possibility
609 that the relative position along the dsDNA molecule dictates where the AT-hook binds, we

610 flipped the relative positions of the two A/T sequences along the dsDNA. For example, we
611 modeled both 5'GGGGAAAAGGGGAAATGGGG 3' and
612 5'GGGGAAATGGGGAAAAGGGG3' sequences (Fig S2G). For each 5W model, the width of
613 the minor groove was measured as the Angstrom distance from the center of the *i* phosphate
614 group on one strand to the *i*+3 phosphate group on the complementary strand (black-dotted lines
615 in Fig S3) using Pymol. 5.8 Angstroms were subtracted from all raw minor groove width
616 measurements to account for the sum of the Van der Waals radii of the two phosphate groups
617 (*68*). An average minor groove width for each dsDNA was calculated from all five AlphaFold3
618 models.
619
620 **TAREAN short-read analysis**
621 100-bp paired-end Illumina reads generated by the Wellcome Sanger Institute as part of
622 the Caroli Genome Project (ENA Run: ERR133993) were downloaded from the European
623 Nucleotide Archive and processed using the Galaxy RepeatExplorer webserver (*69*). Read quality
624 was assessed with FastQC. Using the "Preprocessing of FASTQ paired-end reads" tools in
625 Galaxy, the final base of each read was trimmed due to low average quality (<Q20), reads
626 containing ambiguous nucleotides (N's) were removed, and the remaining read pairs were
627 interleaved into a single FASTA file. This processed and interleaved dataset was used as input for
628 TAREAN (*50*), which analyzed 397,722 reads with cluster merging enabled.
629
630 **Satellite DNA alignment and tree generation**
631 Consensus sequences of *M. caroli* satellites (59-bp, 60-bp and 79-bp) and the sub-repeats
632 of *M. musculus* satellites (minor and major) were aligned using MAFFT using standard settings.
633 The resulting alignment (Fig S4E) had columns with gaps removed and was used to generate a
634 PHYLIP Neighbor Joining Tree using an F84 distance matrix model.
635
636 **Calculating $A_{\geq 4}T$ frequency in satellite-positive *M. caroli* short reads**
637 The same Illumina short-read dataset described above was mapped to concatenated
638 consensus sequences representing either the top 59-bp variants or the 60/79-bp variants using
639 minimap2 with the -sr preset, which is optimized for short-read alignment. Only the best
640 alignment for each read was retained. Reads mapping uniquely to the 59-bp consensus (6,738,028
641 reads), the 60/79-bp consensus (8,952,232 reads), and an equal number of randomly selected
642 reads (8,952,232 reads) were extracted into separate FASTA files. To quantify $A_{\geq 4}T$ motif
643 frequency, we counted all occurrences of the $A_{\geq 4}T$ pattern in each FASTA file using

644 vcountPDict() function in Biostrings package in R and divided this by the total read length in that
645 file.
646
647 **59-bp gRNA design, synthesis and microinjection**

648 To design gRNAs that capture the sequence diversity of the 59-bp satellite arrays, we used
649 the vcountPDict() function in Biostrings to identify the most frequent 20-bp sequences located
650 upstream of a PAM site (NGG) within reads mapping to the 59-bp TAREAN cluster. From these
651 candidates, we selected six abundant gRNAs that do not contain PAM sites in the 60/79-bp
652 satellites: AAATGAGAAAATCACACTGT, AAATGAGAAAATCACACTGC,
653 AAATGAGAAAATCACACAGT, TCTCATTTTTTGACTTTTTC,
654 TCTCATTTTCCACCTTTTTC, TCTCATTTTTCGACTTTTTC (see Fig S4E). gRNAs were
655 synthesized using the GeneArt Precision gRNA Synthesis Kit. A gRNA cocktail (20 ng/μL total)
656 and mRNA encoding dCas9:mCherry (100 ng/μL) were co-injected into MII eggs together with
657 *M. caroli* sperm heads during ICSI. Embryos were arrested in metaphase by treatment with the
658 APC/C inhibitor proTAME (5 μM) and then fixed and stained as described above.
659
660 **Long-read *Mus caroli* genome assembly, evaluation and repeat mapping**

661 Flash-frozen tissue was obtained from the Jackson Laboratory from female *Mus caroli*
662 mice (CAROLI/EiJ, Strain: 000926; RRID:IMSR_JAX:000926). DNA was extracted from tail
663 tissue using the Qiagen MagAttract HMW DNA kit according to the manufacturer's protocol. The
664 Pacbio SEQUEL platform was used to generate HiFi reads according to the manufacturer's
665 protocols. Hi-C sequencing libraries were prepared using the Arima proximity ligation method
666 following the manufacturer's protocols for tissue preparation. Libraries were sequenced on the
667 Illumina NovoSeq platform. A total of 94 Gbp of HiFi, and 430 Gbp of Hi-C sequencing data was
668 generated and used for genome assembly.

669 HiFi reads were assembled into contigs using hifiasm (v0.25.0-r726) (*70*) with
670 duplication-purging disabled, a recommended setting for highly inbred genomes. Scaffolding was
671 performed using YaHS (*71*), which combines contigs into scaffolds by leveraging Hi-C contact
672 frequency patterns. Hi-C reads were aligned to the primary contig assembly using bwa-mem2
673 (*72*) with Hi-C specific flags (-5SP), which improve the handling of chimeric Hi-C junction reads.
674 The resulting alignments were fed into samtools (*73*) for BAM generation and processing. Mate-
675 pair information was corrected using fixmate with the -m flag, reads were sorted by genomic
676 coordinates, and PCR and optical duplicates were removed using markdup with the -r flag. The

677 processed Hi-C BAM and our primary assembly were used as inputs for YaHS, which was run
678 using default parameters.
679 The continuity of the assembled scaffolds was evaluated using assembly-stats. The final
680 scaffolded assembly totaled 2.84 Gb across 1,194 scaffolds with a N50 of 126 Mb and 0.005%
681 gap content (1,393 gaps). Genome completeness (in terms of gene content) was assessed using
682 BUSCO (v6.0.0) (*74*) with the mammalia_odb10 lineage dataset (Fig S5B). Most of the assembly
683 is composed of 20 chromosome-sized scaffolds, which were mapped against a previously
684 published *M. caroli* reference genome (CAROLI_EiJ_v1.1, Ensembl Release 115,
685 GCA_900094665.2) (*75*) using minimap2 (*76*) with asm5 preset parameters. Alignment between
686 the genomes was visualized using the pafr package in R to evaluate large-scale structural
687 accuracy and identify potential scaffolding errors.
688 We identified the genomic locations of 59-bp and 60/79-bp satellite arrays using BLASTN
689 with parameters optimized for short, repetitive sequences (wordsize=7, dust=no). Because the 59-
690 bp satellite family exhibits substantial sequence diversity, we used the top eleven consensus
691 sequences generated by TAREAN's 27-mer analysis as query sequences. In contrast, the 60/79-bp
692 satellite family is highly homogenous, so the single 79-bp consensus sequence produced by
693 TAREAN was used as a query for this repeat class, which also captures the closely related 60-bp
694 variant. Telomere arrays were detected using the matchPattern() function from the Biostrings to
695 find occurrences of four tandem telomere repeats (i.e. TTAGGGTTAGGGTTAGGGTTAGGG).

696

697 **Live imaging of *Mus musculus* zygotes**

698 *Superovulation, collection, and in vitro fertilization of mouse eggs*

699 CF-1 female mice aged 8-12 weeks were superovulated by intraperitoneal injection of 7.5
700 IU pregnant mare serum gonadotropin (PMSG, Medix Biochemica) followed 48 hours later by
701 7.5 IU human chorionic gonadotropin (hCG, Prospec) injection . Male mice were sacrificed and
702 their epididymides were carefully dissected and transferred to a dish containing 100uL of human
703 tubal fluid medium (HTF, MiliporeSigma, MR-070-D), overlaid with mineral oil (FujiFilm
704 Biosciences, 9305). Sperm were collected and incubated in HTF for one hour at 37°C in a
705 humidified atmosphere containing 5% CO_2 for capacitation. Female mice were sacrificed 15-17
706 hours after hCG injection, and their oviducts were dissected and transferred into a fertilization
707 dish containing 200uL HTF overlaid with mineral oil. Oviducts were torn using a 27G
708 hypodermic needle and cumulus oocyte complexes (COCs) were collected and placed in HTF. *In*
709 *vitro* fertilization (IVF) was performed by adding 2–3 µl of capacitated sperm to the fertilization

710 drop. After three hours of IVF culture, zygotes were washed through six drops of K+ Simplex
711 Optimised Medium (KSOM, MiliporeSigma, MR-101).

712 *Cloning, in vitro cRNA synthesis, and microinjection*

713 The H2B:mCherry construct was linearized using Nde I (New England BioLabs). After
714 DNA linearization, the digests were purified (Qiagen, QIAquick PCR Purification) and in vitro
715 transcription was carried out using the mMessage mMachine T7 kit (Ambion) according to the
716 manufacturer's instructions. Finally, the cRNA was purified using an RNAEasy kit (Qiagen) and
717 stored at - 80°C. Zygotes were microinjected 4 hours post-insemination with 5 pl of the cRNA
718 using Eppendorf FemtoJet 4i. Advanced KSOM was used as the microinjection medium. The
719 injected zygotes were allowed to recover for 2 hours at 37°C in a humidified atmosphere
720 containing 5% CO_2 to allow for cRNA expression before live imaging. Microtubules were labeled
721 in live embryos using SiR-tubulin (Cytoskeleton, CY-SC002), added to the culture medium at a
722 final concentration of 100 nM. Microtubule images are not shown in Fig 5E.

723 *Confocal microscopy*

724 For live-cell imaging, zygotes were transferred to small drops of KSOM medium under
725 mineral oil in glass-bottom dishes (MatTek). Dishes were placed in a microenvironmental
726 chamber (UNO-T-H-CO2, Okolab) and kept at 37°C in a humidified atmosphere containing 5%
727 CO_2. Images were acquired using a Leica Stellaris 5 confocal microscope using a 40x oil-
728 immersion objective. Z-stacks were acquired every 8 min covering a 70 µm range (5 µm step
729 size) for a total duration of 8 hours.

730 *Data Analysis*

731 Zygotes were analyzed using the National Institutes of Health (NIH) ImageJ software
732 (NIH, Bethesda, MD, USA) and Imaris software (Oxford Instruments, Buckinghamshire, UK).
733 Paternal and maternal chromosomes were distinguished based on the larger size of the paternal
734 pronucleus (Odor & Blandua, 1951). 3D reconstruction and manual object classification in Imaris
735 was used to differentially label paternal and maternal chromosomes at nuclear envelope
736 breakdown, and chromosomes were then automatically tracked using Imaris. Lagging
737 chromosomes were identified as chromosomes remaining near the spindle equator after anaphase
738 onset.
739
740
741

References

1. N. Altemose, G. A. Logsdon, A. V. Bzikadze, P. Sidhwani, S. A. Langley, G. V. Caldas, S. J. Hoyt, L. Uralsky, F. D. Ryabov, C. J. Shew, M. E. G. Sauria, M. Borchers, A. Gershman, A. Mikheenko, V. A. Shepelev, T. Dvorkina, O. Kunyavskaya, M. R. Vollger, A. Rhie, A. M. McCartney, M. Asri, R. Lorig-Roach, K. Shafin, J. K. Lucas, S. Aganezov, D. Olson, L. G. de Lima, T. Potapova, G. A. Hartley, M. Haukness, P. Kerpedjiev, F. Gusev, K. Tigyi, S. Brooks, A. Young, S. Nurk, S. Koren, S. R. Salama, B. Paten, E. I. Rogaev, A. Streets, G. H. Karpen, A. F. Dernburg, B. A. Sullivan, A. F. Straight, T. J. Wheeler, J. L. Gerton, E. E. Eichler, A. M. Phillippy, W. Timp, M. Y. Dennis, R. J. O'Neill, J. M. Zook, M. C. Schatz, P. A. Pevzner, M. Diekhans, C. H. Langley, I. A. Alexandrov, K. H. Miga, Complete genomic and epigenetic maps of human centromeres. *Science* **376**, eabl4178 (2022).

2. D. Dudka, J. M. Dawicki-McKenna, X. Sun, K. Beeravolu, T. Akera, M. A. Lampson, B. E. Black, Satellite DNA shapes dictate pericentromere packaging in female meiosis. *Nature* **638**, 814–822 (2025).

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

4. D. P. Melters, K. R. Bradnam, H. A. Young, N. Telis, M. R. May, J. Ruby, R. Sebra, P. Peluso, J. Eid, D. Rank, J. Garcia, J. L. DeRisi, T. Smith, C. Tobias, J. Ross-Ibarra, I. Korf, S. W. Chan, Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**, R10 (2013).

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

6. U. P. Arora, B. A. Sullivan, B. L. Dumont, Variation in the CENP-A sequence association landscape across diverse inbred mouse strains. *Cell Reports* **42**, 113178 (2023).

7. S. Nurk, S. Koren, A. Rhie, M. Rautiainen, A. V. Bzikadze, A. Mikheenko, M. R. Vollger, N. Altemose, L. Uralsky, A. Gershman, S. Aganezov, S. J. Hoyt, M. Diekhans, G. A. Logsdon, M. Alonge, S. E. Antonarakis, M. Borchers, G. G. Bouffard, S. Y. Brooks, G. V. Caldas, N.-C. Chen, H. Cheng, C.-S. Chin, W. Chow, L. G. de Lima, P. C. Dishuck, R. Durbin, T. Dvorkina, I. T. Fiddes, G. Formenti, R. S. Fulton, A. Fungtammasan, E. Garrison, P. G. S. Grady, T. A. Graves-Lindsay, I. M. Hall, N. F. Hansen, G. A. Hartley, M. Haukness, K. Howe, M. W. Hunkapiller, C. Jain, M. Jain, E. D. Jarvis, P. Kerpedjiev, M. Kirsche, M. Kolmogorov, J. Korlach, M. Kremitzki, H. Li, V. V. Maduro, T. Marschall, A. M. McCartney, J. McDaniel, D. E. Miller, J. C. Mullikin, E. W. Myers, N. D. Olson, B. Paten, P. Peluso, P. A. Pevzner, D. Porubsky, T. Potapova, E. I. Rogaev, J. A. Rosenfeld, S. L. Salzberg, V. A. Schneider, F. J. Sedlazeck, K. Shafin, C. J. Shew, A. Shumate, Y. Sims, A. F. A. Smit, D. C. Soto, I. Sović, J. M. Storer, A. Streets, B. A. Sullivan, F. Thibaud-Nissen, J. Torrance, J. Wagner, B. P. Walenz, A. Wenger, J. M. D. Wood, C. Xiao, S. M. Yan, A. C. Young, S. Zarate, U. Surti, R. C. McCoy, M. Y. Dennis, I. A. Alexandrov, J. L. Gerton, R. J. O'Neill, W. Timp, J. M. Zook, M. C. Schatz, E. E. Eichler, K. H. Miga, A. M. Phillippy, The complete sequence of a human genome. *Science* **376**, 44–53 (2022).

8. G. A. Logsdon, E. E. Eichler, The Dynamic Structure and Rapid Evolution of Human Centromeric Satellite DNA. *Genes* **14**, 92 (2022).

9. C. W. Gambogi, N. Pandey, J. M. Dawicki-McKenna, U. P. Arora, M. A. Liskovykh, J. Ma, P. Lamelza, V. Larionov, M. A. Lampson, G. A. Logsdon, B. L. Dumont, B. E. Black, Centromere innovations within a mouse species. *Science Advances* **9**, eadi5764 (2023).

10. J. Packiaraj, J. Thakur, DNA satellite and chromatin organization at mouse centromeres and pericentromeres. *Genome Biol* **25**, 52 (2024).

11. S. Gao, K. K. Oshima, S.-C. Chuang, M. Loftus, A. Montanari, D. S. Gordon, H. G. S. V. Consortium, H. P. R. Consortium, P. Hsieh, M. K. Konkel, M. Ventura, G. A. Logsdon, A global view of human centromere variation and evolution. bioRxiv [Preprint] (2025). https://doi.org/10.64898/2025.12.09.693231.

12. S. Henikoff, The Centromere Paradox: Stable Inheritance with Rapidly Evolving DNA. *Science* **293**, 1098–1102 (2001).

13. S. Fioriniello, D. Marano, F. Fiorillo, M. D'Esposito, F. Della Ragione, Epigenetic Factors that Control Pericentric Heterochromatin Organization in Mammals. *Genes* **11**, 595 (2020).

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

15. M. Carmena, M. Wheelock, H. Funabiki, W. C. Earnshaw, The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat Rev Mol Cell Biol* **13**, 789–803 (2012).

16. S. A. Kawashima, T. Tsukahara, M. Langegger, S. Hauf, T. S. Kitajima, Y. Watanabe, Shugoshin enables tension-generating attachment of kinetochores by loading Aurora to centromeres. *Genes Dev.* **21**, 420–435 (2007).

17. A. Janssen, S. U. Colmenares, G. H. Karpen, Heterochromatin: Guardian of the Genome. *Annu. Rev. Cell Dev. Biol.* **34**, 265–288 (2018).

18. C.-J. Lin, F. M. Koh, P. Wong, M. Conti, M. Ramalho-Santos, Hira-Mediated H3.3 Incorporation Is Required for DNA Replication and Ribosomal RNA Transcription in the Mouse Zygote. *Developmental Cell* **30**, 268–279 (2014).

19. B. Loppin, E. Bonnefoy, C. Anselme, A. Laurençon, T. L. Karr, P. Couble, The histone H3.3 chaperone HIRA is essential for chromatin assembly in the male pronucleus. *Nature* **437**, 1386–1390 (2005).

20. M. Puschendorf, R. Terranova, E. Boutsma, X. Mao, K. Isono, U. Brykczynska, C. Kolb, A. P. Otte, H. Koseki, S. H. Orkin, M. van Lohuizen, A. H. F. M. Peters, PRC1 and Suv39h specify parental asymmetry at constitutive heterochromatin in early mouse embryos. *Nat Genet* **40**, 411–420 (2008).

21. N. Saksouk, T. K. Barth, C. Ziegler-Birling, N. Olova, A. Nowak, E. Rey, J. Mateos-Langerak, S. Urbach, W. Reik, M.-E. Torres-Padilla, A. Imhof, J. Déjardin, Redundant Mechanisms to Form Silent Chromatin at Pericentromeric Regions Rely on BEND3 and DNA Methylation. *Molecular Cell* **56**, 580–594 (2014).

22. M. Tardat, M. Albert, R. Kunzmann, Z. Liu, L. Kaustov, R. Thierry, S. Duan, U. Brykczynska, C. H. Arrowsmith, A. H. F. M. Peters, Cbx2 Targets PRC1 to Constitutive

823 Heterochromatin in Mouse Zygotes in a Parent-of-Origin-Dependent Manner. *Molecular*
824 *Cell* **58**, 157–171 (2015).

825 23. A. Santenard, C. Ziegler-Birling, M. Koch, L. Tora, A. J. Bannister, M.-E. Torres-Padilla,
826 Heterochromatin formation in the mouse embryo requires critical residues of the histone
827 variant H3.3. *Nat Cell Biol* **12**, 853–862 (2010).

828 24. Z. Liu, M. Tardat, M. E. Gill, H. Royo, R. Thierry, E. A. Ozonov, A. H. Peters, SUMO
829 ylated PRC 1 controls histone H3.3 deposition and genome integrity of embryonic
830 heterochromatin. *EMBO J* **39** (2020).

831 25. T. Kawaguchi, S. Machida, H. Kurumizaka, H. Tagami, J. Nakayama, Phosphorylation of
832 CBX2 controls its nucleosome-binding specificity. *The Journal of Biochemistry* **162**, 343–
833 355 (2017).

834 26. S. Battista, M. Fedele, L. Secco, A. M. D. Ingo, R. Sgarra, G. Manfioletti, Binding to the
835 Other Side: The AT-Hook DNA-Binding Domain Allows Nuclear Factors to Exploit the
836 DNA Minor Groove. *IJMS* **25**, 8863 (2024).

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

840 28. J. R. Huth, C. A. Bewley, M. S. Nissen, J. N. S. Evans, R. Reeves, A. M. Gronenborn, G. M.
841 Clore, The solution structure of an HMG-I(Y)–DNA complex defines a new architectural
842 minor groove binding motif. *Nat Struct Mol Biol* **4**, 657–665 (1997).

843 29. R. Reeves, Structure and Function of the HMGI(Y) Family of Architectural Transcription
844 Factors. *Environmental Health Perspectives* **108**, 803–809 (2000).

845 30. P. Stein, R. M. Schultz, "Chapter 14 - ICSI in the Mouse" in *Methods in Enzymology*, P. M.
846 Wassarman, P. M. Soriano, Eds. (Academic Press, 2010;
847 https://www.sciencedirect.com/science/article/pii/S0076687910760146)vol. 476 of *Guide to*
848 *Techniques in Mouse Development, Part A: Mice, Embryos, and Cells, 2nd Edition*, pp.
849 251–262.

850 31. D. Jukam, S. A. M. Shariati, J. M. Skotheim, Zygotic Genome Activation in Vertebrates.
851 *Developmental Cell* **42**, 316–332 (2017).

852 32. P. M. Ferree, D. A. Barbash, Species-Specific Heterochromatin Prevents Mitotic
853 Chromosome Segregation to Cause Hybrid Lethality in Drosophila. *PLOS Biology* **7**,
854 e1000234 (2009).

855 33. C. Josefsson, B. Dilkes, L. Comai, Parent-Dependent Loss of Gene Silencing during
856 Interspecies Hybridization. *Current Biology* **16**, 1322–1328 (2006).

857 34. R. Gibeaux, R. Acker, M. Kitaoka, G. Georgiou, I. van Kruijsbergen, B. Ford, E. M.
858 Marcotte, D. K. Nomura, T. Kwon, G. J. C. Veenstra, R. Heald, Paternal chromosome loss
859 and metabolic crisis contribute to hybrid inviability in Xenopus. *Nature* **553**, 337–341
860 (2018).

861 35. L. Wang, J. L. Brown, R. Cao, Y. Zhang, J. A. Kassis, R. S. Jones, Hierarchical Recruitment
862 of Polycomb Group Silencing Complexes. *Molecular Cell* **14**, 637–646 (2004).

863 36. R. Cao, L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage, P. Tempst, R. S. Jones, Y.
864 Zhang, Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science*
865 **298**, 1039–1043 (2002).

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

868 38. H. Goto, H. Kosako, Y. Tomono, K. Ajiro, M. Fujita, M. Sakurai, K. Okawa, A. Iwamatsu,
869 T. Okigaki, T. Takahashi, M. Inagaki, Identification of a Novel Phosphorylation Site on
870 Histone H3 Coupled with Mitotic Chromosome Condensation. *Journal of Biological*
871 *Chemistry* **274**, 25543–25549 (1999).

872 39. W. Fischle, B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia, J. Shabanowitz,
873 D. F. Hunt, H. Funabiki, C. D. Allis, Regulation of HP1–chromatin binding by histone H3
874 methylation and phosphorylation. *Nature* **438**, 1116–1122 (2005).

875 40. T. Hirota, J. J. Lipp, B.-H. Toh, J.-M. Peters, Histone H3 serine 10 phosphorylation by
876 Aurora B causes HP1 dissociation from heterochromatin. *Nature* **438**, 1176–1180 (2005).

877 41. S. Y. Breusegem, R. M. Clegg, F. G. Loontiens, Base-sequence specificity of Hoechst 33258
878 and DAPI binding to five (A/T)4 DNA sites with kinetic evidence for more than one high-
879 affinity Hoechst 33258-AATT complex. *Journal of Molecular Biology* **315**, 1049–1061
880 (2002).

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

883 43. F. A. Tanious, J. M. Veal, H. Buczak, L. S. Ratmeyer, W. D. Wilson, DAPI (4',6-diamidino-
884 2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and
885 intercalation at AU sites. *Biochemistry* **31**, 3103–3112 (1992).

886 44. W. Traut, D. Weichenhan, U. Eickhoff, H. Winking, Origin of the Chromosome 1 HSR of
887 the House Mouse Detected by CGH. *Chromosome Res* **7**, 649–653 (1999).

888 45. S. Thakur, D. I. Cattoni, M. Nöllmann, The fluorescence properties and binding mechanism
889 of SYTOX green, a bright, low photo-damage DNA intercalating agent. *Eur Biophys J* **44**,
890 337–348 (2015).

891 46. R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, B. Honig, The role of DNA shape in
892 protein–DNA recognition. *Nature* **461**, 1248–1253 (2009).

893 47. J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L.
894 Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M.
895 O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O.
896 Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M.
897 Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A.
898 Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D.
899 Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper,

900 Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**,
901 493–500 (2024).

902 48. M. J. Solomon, F. Strauss, A. Varshavsky, A mammalian high mobility group protein
903 recognizes any stretch of six A.T base pairs in duplex DNA. *Proceedings of the National*
904 *Academy of Sciences* **83**, 1276–1280 (1986).

905 49. D. Kipling, A. R. Mitchell, H. Masumoto, H. E. Wilson, L. Nicol, H. J. Cooke, CENP-B
906 binds a novel centromeric sequence in the Asian mouse Mus caroli. *Mol Cell Biol* **15**, 4009–
907 4020 (1995).

908 50. P. Novák, L. Ávila Robledillo, A. Koblížková, I. Vrbová, P. Neumann, J. Macas, TAREAN:
909 a computational tool for identification and characterization of satellite DNA from
910 unassembled short reads. *Nucleic Acids Res* **45**, e111 (2017).

911 51. P. E. Warburton, H. F. Willard, Genomic analysis of sequence variation in tandemly
912 repeated DNA: Evidence for localized homogeneous sequence domains within arrays of α-
913 satellite DNA. *Journal of Molecular Biology* **216**, 3–16 (1990).

914 52. H. Aihara, T. Nakagawa, H. Mizusaki, M. Yoneda, M. Kato, M. Doiguchi, Y. Imamura, M.
915 Higashi, T. Ikura, T. Hayashi, Y. Kodama, M. Oki, T. Nakayama, E. Cheung, H. Aburatani,
916 K. Takayama, H. Koseki, S. Inoue, Y. Takeshima, T. Ito, Histone H2A T120
917 Phosphorylation Promotes Oncogenic Transformation via Upregulation of Cyclin D1.
918 *Molecular Cell* **64**, 176–188 (2016).

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

921 54. C. S. Schober, F. Aydiner, C. J. Booth, E. Seli, V. Reinke, The kinase VRK1 is required for
922 normal meiotic progression in mammalian oogenesis. *Mechanisms of Development* **128**,
923 178–190 (2011).

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

926 56. J. G. DeLuca, W. E. Gall, C. Ciferri, D. Cimini, A. Musacchio, E. D. Salmon, Kinetochore
927 Microtubule Dynamics and Attachment Stability Are Regulated by Hec1. *Cell* **127**, 969–982
928 (2006).

929 57. T. S. Kitajima, T. Sakuno, K. Ishiguro, S. Iemura, T. Natsume, S. A. Kawashima, Y.
930 Watanabe, Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature*
931 **441**, 46–52 (2006).

932 58. C. G. Riedel, V. L. Katis, Y. Katou, S. Mori, T. Itoh, W. Helmhart, M. Gálová, M.
933 Petronczki, J. Gregan, B. Cetin, I. Mudrak, E. Ogris, K. Mechtler, L. Pelletier, F. Buchholz,
934 K. Shirahige, K. Nasmyth, Protein phosphatase 2A protects centromeric sister chromatid
935 cohesion during meiosis I. *Nature* **441**, 53–61 (2006).

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

938 60. D. Cimini, B. Howell, P. Maddox, A. Khodjakov, F. Degrassi, E. D. Salmon, Merotelic
939 Kinetochore Orientation Is a Major Mechanism of Aneuploidy in Mitotic Mammalian Tissue
940 Cells. *J Cell Biol* **153**, 517–528 (2001).

941 61. G. A. Logsdon, A. N. Rozanski, F. Ryabov, T. Potapova, V. A. Shepelev, C. R. Catacchio,
942 D. Porubsky, Y. Mao, D. Yoo, M. Rautiainen, S. Koren, S. Nurk, J. K. Lucas, K. Hoekzema,
943 K. M. Munson, J. L. Gerton, A. M. Phillippy, M. Ventura, I. A. Alexandrov, E. E. Eichler,
944 The variation and evolution of complete human centromeres. *Nature* **629**, 136–145 (2024).

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

947 63. Y. Rong, Y.-Z. Zhu, J. Yu, Y.-W. Wu, S.-Y. Ji, Y. Zhou, Y. Jiang, J. Jin, H.-Y. Fan, L.
948 Shen, Q.-Q. Sha, USP16-mediated histone H2A lysine-119 deubiquitination during oocyte
949 maturation is a prerequisite for zygotic genome activation. *Nucleic Acids Research* **50**,
950 5599–5616 (2022).

951 64. Q. Duan, H. Chen, M. Costa, W. Dai, Phosphorylation of H3S10 Blocks the Access of
952 H3K9 by Specific Antibodies and Histone Methyltransferase: IMPLICATION IN
953 REGULATING CHROMATIN DYNAMICS AND EPIGENETIC INHERITANCE
954 DURING MITOSIS*. *Journal of Biological Chemistry* **283**, 33585–33590 (2008).

955 65. A. Das, A. Iwata-Otsubo, A. Destouni, J. M. Dawicki-McKenna, K. G. Boese, B. E. Black,
956 M. A. Lampson, Epigenetic, genetic and maternal effects enable stable centromere
957 inheritance. *Nat Cell Biol* **24**, 748–756 (2022).

958 66. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S.
959 Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K.
960 Eliceiri, P. Tomancak, A. Cardona, Fiji: an open-source platform for biological-image
961 analysis. *Nat Methods* **9**, 676–682 (2012).

962 67. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan,
963 V. J. Carey, Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol*
964 **9**, e1003118 (2013).

965 68. E. Gavathiotis, G. J. Sharman, M. S. Searle, Sequence-dependent variation in DNA minor
966 groove width dictates orientational preference of Hoechst 33258 in A-tract recognition:
967 solution NMR structure of the 2:1 complex with d(CTTTTGCAAAAG)2. *Nucleic Acids Res*
968 **28**, 728–735 (2000).

969 69. P. Novák, P. Neumann, J. Pech, J. Steinhaisl, J. Macas, RepeatExplorer: a Galaxy-based web
970 server for genome-wide characterization of eukaryotic repetitive elements from next-
971 generation sequence reads. *Bioinformatics* **29**, 792–793 (2013).

972 70. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo
973 assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).

974 71. C. Zhou, S. A. McCarthy, R. Durbin, YaHS: yet another Hi-C scaffolding tool.
975 *Bioinformatics* **39**, btac808 (2023).

976 72. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform.
977 *Bioinformatics* **25**, 1754–1760 (2009).

73. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

74. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

75. D. Thybert, M. Roller, F. C. P. Navarro, I. Fiddes, I. Streeter, C. Feig, D. Martin-Galvez, M. Kolmogorov, V. Janoušek, W. Akanni, B. Aken, S. Aldridge, V. Chakrapani, W. Chow, L. Clarke, C. Cummins, A. Doran, M. Dunn, L. Goodstadt, K. Howe, M. Howell, A.-A. Josselin, R. C. Karn, C. M. Laukaitis, L. Jingtao, F. Martin, M. Muffato, S. Nachtweide, M. A. Quail, C. Sisu, M. Stanke, K. Stefflova, C. Van Oosterhout, F. Veyrunes, B. Ward, F. Yang, G. Yazdanifar, A. Zadissa, D. J. Adams, A. Brazma, M. Gerstein, B. Paten, S. Pham, T. M. Keane, D. T. Odom, P. Flicek, Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Res.* **28**, 448–459 (2018).

76. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

Acknowledgments

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. We thank Dr. Beth L. Dumont for sending us PAHARI/EiJ mice from Jackson Laboratory. We thank Emma Hamlin for help with data analysis. We thank Drs. Mia Levine, Ben Black and Richard Schultz for helpful discussions and comments on the manuscript.

Funding:

Helen Hay Whitney Foundation (PL)

National Institutes of Health grant 1K99GM152835 (PL)

National Institutes of Health grant R35GM122475 (MAL)

National Institutes of Health grant R35GM142537 (AZB)

National Institutes of Health grant R01HD113358 (AZB)

Author contributions:

Conceptualization: PL, MAL

Methodology: PL, MP, KL

1015 Investigation: PL, MP, KL
1016 Visualization: PL, MP, KL
1017 Supervision: MAL, AZB, TMK
1018 Writing—original draft: PL, MAL
1019 Writing—review & editing: PL, KL, MAL, AZB, TMK
1020
1021 **Competing interests:** Authors declare that they have no competing interests.
1022
1023 **Data and materials availability:** All data needed to evaluate the conclusions in the paper are
1024 present in the paper and/or the Supplementary Materials.
1025
1026
1027 **Supplementary Materials Table of Contents**
1028
1029 **Fig. S1.** Paternal *M. pahari* pericentromeres recruit PRC1 during zygotic G2.
1030 **Fig. S2.** Satellite sequence determinants of PRC1 binding.
1031 **Fig. S3.** AlphaFold3 A/T-run minor groove width measurements.
1032 **Fig. S4.** The 59-bp *M. caroli* satellite is diverse and similar to major satellite.
1033 **Fig. S5.** Quality assessment and repeat profiling of the long-read *M. caroli* assembly.
1034 **Fig. S6.** Reduced CPC at paternal *M. caroli* pericentromeres.
1035 **Fig. S7.** Conservation of the CBX2 AT-hook across Muridae.
1036 **Table S1.** Summary of *in silico* competitive binding assays.
1037 **Table S2.** Quantification of A_nT frequency across satellite arrays and chromosome arms in *Mus*
1038 species.
1039 **Table S3.** 60/79-bp satellite k-mer report.
1040 **Table S4.** 59-bp satellite k-mer report.
1041 **Movie S1.** Live imaging of *M. musculus* zygotic mitosis.
1042 **Movie S2.** 3D reconstruction of *M. musculus* zygotic mitosis.
1043
1044 **Figure Captions**
1045
1046 **Fig 1. Divergence in pericentric H2AK119ub1 heterochromatin and PRC1 recruitment. (A)**
1047 Schematic showing PRC1 recruitment to DNA via the AT-hook of its CBX2 subunit, which binds
1048 into the minor groove of A/T runs present in satellite DNA. Chromodomain labelled as “CD”. **(B)**

1049 Mouse zygotes are generated by injecting a *M. musculus* egg arrested in meiosis II with the sperm
1050 of mouse species with divergent pericentric satellite DNAs. Previously characterized satellites
1051 from each species are listed. **(C-E)** Zygotes generated with sperm from the indicated species were
1052 arrested in mitosis with a kinesin-5 inhibitor (STLC, C) or by APC/C inhibition (proTAME, D),
1053 then fixed and stained for H2AK119ub1 or RNF2 (green), H3K9me3 (magenta) to mark maternal
1054 *M. musculus* chromosomes, and DAPI (gray). Arrowheads (D) highlight the slight enrichment of
1055 RNF2 near *M. pahari* centromeres. Graphs (E) show average H2AK119ub1 (n=103-164, *N*=2)
1056 and RNF2 (n=152-160, *N*=2) intensities along *M. musculus* (red), *M. caroli* (yellow) and *M.*
1057 *pahari* (blue) paternal chromosomes, starting from pericentric ends. S.E.M. is indicated by light
1058 band surrounding the mean line. Both *M. musculus* and *M. caroli* chromosomes have significantly
1059 higher H2AK119ub1 intensity than *M. pahari* chromosomes within the first two microns (gray
1060 box), (*** P <0.001). Statistical significance was calculated by a Kruskal-Wallis test, followed by
1061 Dunn's post hoc test with Bonferroni correction. Images are max intensity z-projections, scale
1062 bars 10 μm or 1 μm (insets).
1063
1064 **Fig 2. Paternal pericentromeres that form PRC1 heterochromatin are enriched for narrow**
1065 **DNA minor grooves**. **(A-D)** Zygotes generated with *M. musculus* or *M. pahari* sperm were
1066 arrested in mitosis with a kinesin-5 inhibitor (STLC), then fixed and stained as indicated. Survivin
1067 (CPC subunit) marks pericentromeres, H3K9me3 marks maternal chromosomes, H3K27me3
1068 marks paternal pericentromeres, and DAPI or Sytox Green label DNA. Arrowheads in insets point
1069 to paternal pericentromeres as indicated by the presence of Survivin and absence of H3K9me3
1070 **(A)** or enrichment of H3K27me3 **(C)**. Images are max intensity z-projections, scale bars 10 μm or
1071 1 μm (insets). Graphs show average DAPI (**B**, n=72-165, *N*=2) or Sytox Green (**D**, n=44-66, *N*=2)
1072 intensity along *M. musculus* (red), *M. caroli* (yellow) and *M. pahari* (blue) paternal
1073 chromosomes, starting from pericentric ends. Both *M. musculus* and *M. caroli* chromosomes have
1074 significantly higher DAPI intensity than *M. pahari* chromosomes within the first two microns (B,
1075 gray box), (*** P <0.001). Statistical significance was calculated by a Kruskal-Wallis test,
1076 followed by Dunn's post hoc test with Bonferroni correction.
1077
1078 **Fig 3. *M. musculus* major satellite arrays are enriched for long A_nT sequences that**
1079 **preferentially bind the CBX2 AT-hook**. **(A)** Histograms show the number of various A/T run
1080 lengths per 10 kb bin along portions of two genomic scaffolds containing arrays of either *M.*
1081 *musculus* major satellite (left) or *M. pahari* pi satellite (right), indicated by tandem arrows. **(B)**
1082 The average number per 10 kb bin of all possible 4W (left) and 5W (right) sequences within the

1083 major and pi satellites arrays shown in A. **(C)** Schematic (left) shows *in silico* competition assays
1084 between different A/T run sequences for the binding of a single peptide encoding the AT-hook of
1085 CBX2, using AlphaFold3. Results (right) show AT-hook binding preferences. **(D)** Histograms
1086 show the number of various A_nT sequences per 10 kb bin along the same genomic scaffolds as in
1087 A. **(E)** Top: schematic showing that *M. musculus* and *M. spretus* centromeres/pericentromeres
1088 have inverse proportions of minor (pink) and major (red) satellites. Bottom: consensus minor
1089 satellite sequence with A_nT-sequences in bold and predicted A_nT frequencies. **(F)** Zygotes
1090 generated with *M. musculus* (top row) or *M. spretus* (bottom row) sperm were arrested in mitosis
1091 by APC/C inhibition (proTAME), then fixed and stained for RNF2 (green), H3K9me3 (magenta)
1092 to mark maternal *M. musculus* chromosomes, and DAPI (gray). Images are max intensity z-
1093 projections, scale bars 10 μm or 1 μm (insets). **(G)** Graphs show average RNF2 intensity along *M.*
1094 *musculus* (red, n=166) and *M. spretus* (pink, n=119) paternal chromosomes, starting from
1095 pericentric ends. S.E.M. is indicated by light band surrounding the mean line. *M. musculus*
1096 chromosomes have significantly higher RNF2 intensity than *M. spretus* chromosomes within the
1097 first two microns (gray box), (*** P <0.001) (N=2). Statistical significance was calculated by a
1098 Kruskal-Wallis test, followed by Dunn's post hoc test with Bonferroni correction.
1099

1100 **Fig 4. Arrays of 59-bp A_nT-rich satellites are present at *M. caroli* pericentromeres. (A)**
1101 Consensus sequences and predicted A_nT frequencies of two TAREAN read clusters predicted to
1102 be satellite arrays in *M. caroli*. **(B)** $A_{\geq 4}T$ frequency of *M. caroli* short reads mapped to 59-bp
1103 consensus sequences (6,738,028 reads), 60/79-bp consensus sequences (8,952,232 reads), or
1104 random reads (8,952,232 reads). **(C)** Zygotes generated with *M. caroli* sperm expressing
1105 dCas9:mCherry (green) with gRNAs targeting the 59-bp satellite. Embryos were arrested in
1106 mitosis by APC/C inhibition (proTAME), then fixed and stained for H3K9me3 (magenta) to mark
1107 maternal *M. musculus* chromosomes and DAPI (gray). Representative image is a max intensity z-
1108 projection, scale bars 10 μm or 1 μm (inset) (N=2). **(D)** Schematic showing the location of 59-bp
1109 and telomere repeat arrays on the 20 chromosomal scaffolds of our *M. caroli* long-read assembly.
1110 Repeat array sizes are not to scale. **(E)** Histogram showing the number of various A_nT sequences
1111 per 10 kb bin along the first megabase of *M. caroli* chromosome 16. The position of the 59-bp
1112 array is indicated by the horizontal blue bar. **(F)** The average number of various A_nT sequences
1113 per 10 kb bin in major, 59-bp and pi satellite arrays as well as *M. caroli* chromosome arms.
1114

1115 **Fig 5. PRC1 pericentric heterochromatin inhibits CPC recruitment. (A)** Schematic shows
1116 how H2AK119ub1 can inhibit H2AT121phos and CPC recruitment, resulting in increased

1117 distance between metaphase sister kinetochores at paternal *M. musculus* pericentromeres (right)
1118 vs *M. pahari* pericentromeres (left). (**B)** Pure *M. musculus* zygotes were arrested at metaphase by
1119 APC/C inhibition (proTAME), then fixed and stained for RNF2 (PRC1 subunit, magenta) and
1120 H2AT121phos (green). Top and bottom insets show paternal (RNF2 positive) and maternal
1121 (RNF2 negative) chromosomes, respectively; arrowheads point to pericentromeric ends of
1122 chromosomes. Graphs show average H2AT121phos and RNF2 intensity along maternal (n=56)
1123 and paternal (n=54) chromosomes, starting from pericentric ends. Statistical differences between
1124 maternal and paternal chromosomes from 0-1 and 2-3 microns along their length (gray boxes)
1125 were calculated (*N*=2). **(C)** Zygotes generated with *M. musculus* or *M. pahari* sperm were arrested
1126 in mitosis with a kinesin-5 inhibitor (STLC), then fixed and stained for Survivin (green) and
1127 H3K9me3 (magenta). Arrowheads in insets point to pericentric Survivin staining. Graph shows
1128 the mean Survivin intensity of maternal and paternal pericentromeres in both types of zygotes.
1129 Each point represents a single pericentromere (n=253-374 for each group, *N*=3), and boxes
1130 represent interquartile ranges. **(D)** Zygotes generated with *M. musculus* or *M. pahari* sperm were
1131 arrested at metaphase by APC/C inhibition (proTAME), then fixed and stained for CENP-C
1132 (green) and H3K9me3 (magenta). Arrowhead points to a pair of paternal *M. musculus* sister
1133 kinetochores. Graph shows distances between sister kinetochores for maternal and paternal
1134 chromosomes in both types of zygotes. Each point represents a single pair (n=230-284 for each
1135 group, *N*=3). *P*-values were calculated by a Kruskal-Wallis test followed by Dunn's test using
1136 Bonferroni correction for multiple testing (*** $P < 0.001$). Images are max intensity z-
1137 projections; scale bars 10 μm or 1 μm (insets). **(E)** Live-imaging snapshots show a *M.*
1138 *musculus*♀/*M. musculus*♂ zygote expressing H2B:mCherry (cyan) progressing through the first
1139 mitosis (top panels), with arrowheads indicating lagging chromosomes in anaphase, and a 3D
1140 reconstruction of maternal (magenta) and paternal (blue) chromosomes (bottom panels). Plots
1141 show the fraction of zygotes with lagging chromosomes in anaphase (top) and the proportion of
1142 maternal and paternal chromosomes contributing to these lagging events (bottom). Raw numbers
1143 are shown within each bar (*N*=3). *P*-value was calculated using a one-sided binomial test to assess
1144 whether paternal chromosomes lagged more often than the null expectation of 50% (* $P < 0.05$).
1145

Figure 1

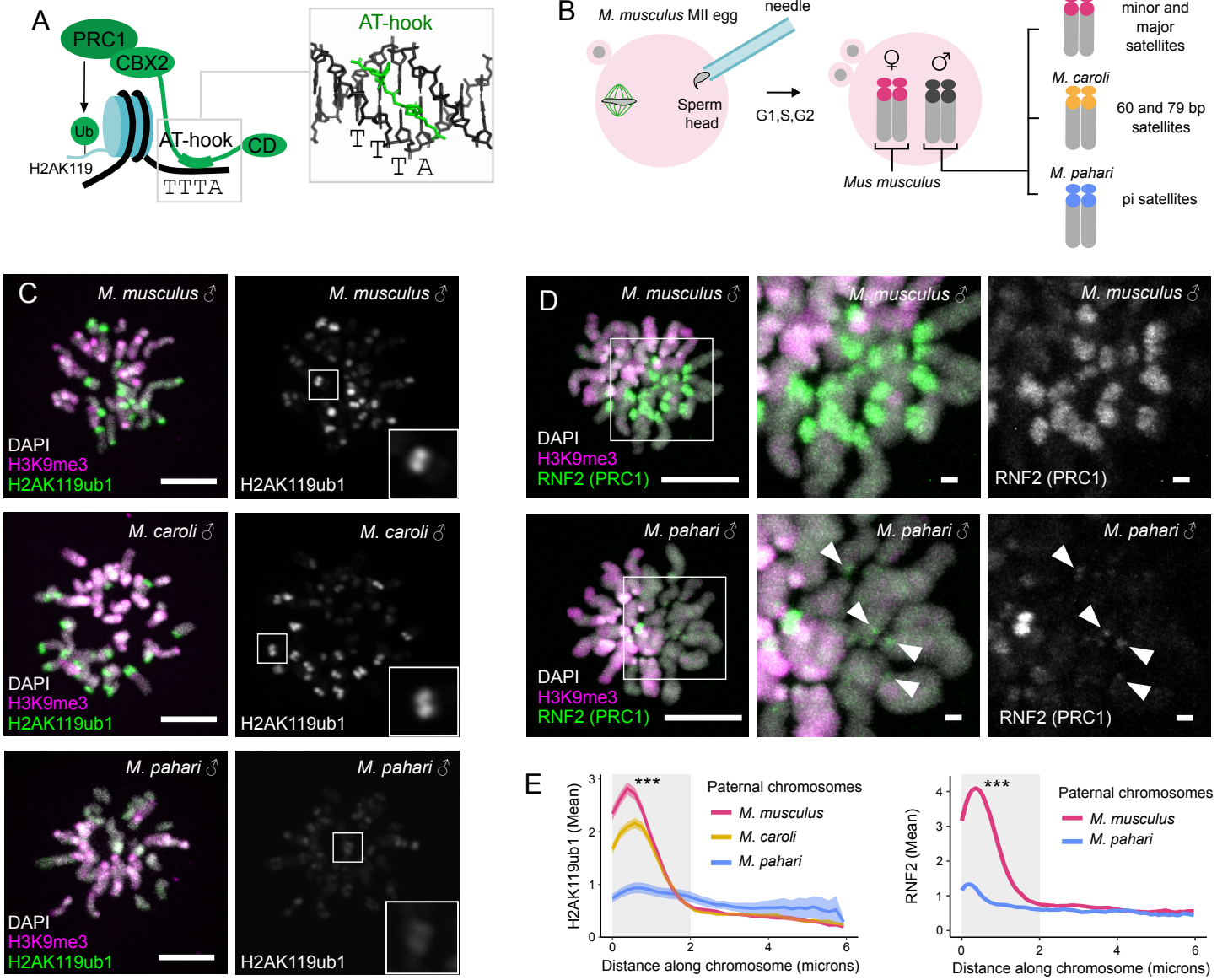

A
PRC1
CBX2
Ub
H2AK119
AT-hook
CD
TTTA
AT-hook
B
M. musculus MII egg
ICSI needle
Sperm head
G1,S,G2
Mus musculus
M. musculus
minor and major satellites
M. caroli
60 and 79 bp satellites
M. pahari
pi satellites
C
M. musculus ♂
DAPI
H3K9me3
H2AK119ub1
M. caroli ♂
M. pahari ♂
D
RNF2 (PRC1)
E
H2AK119ub1 (Mean)
Paternal chromosomes
M. musculus
M. caroli
M. pahari
***
Distance along chromosome (microns)
RNF2 (Mean)

Figure 2

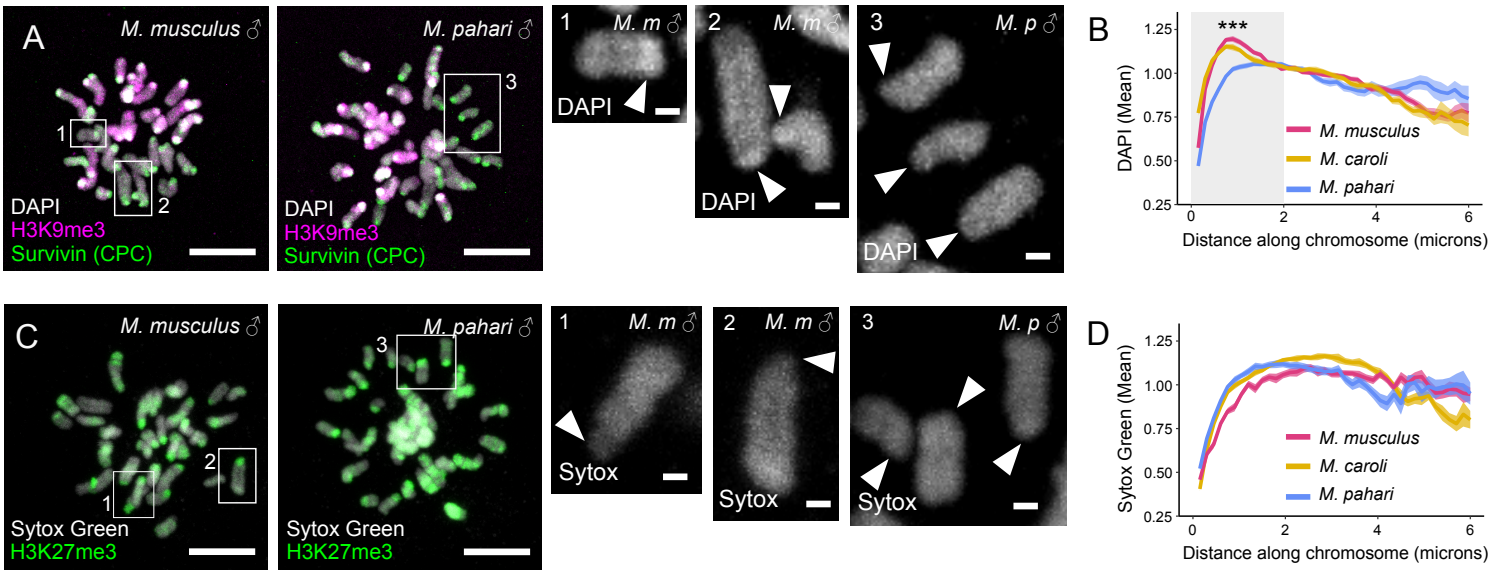
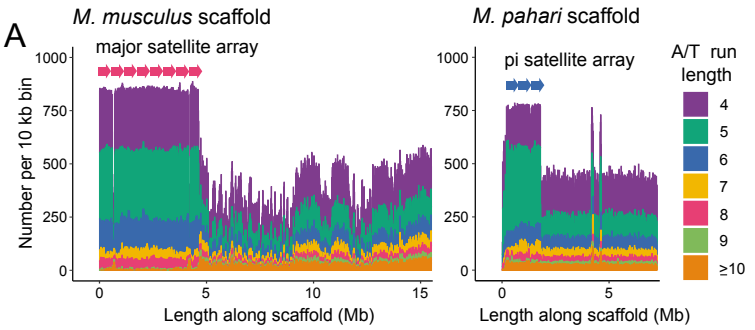
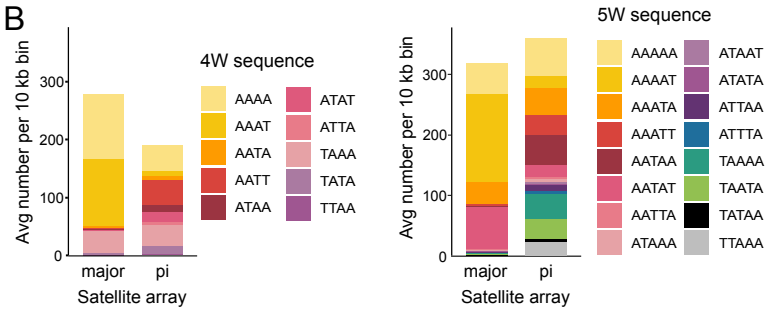
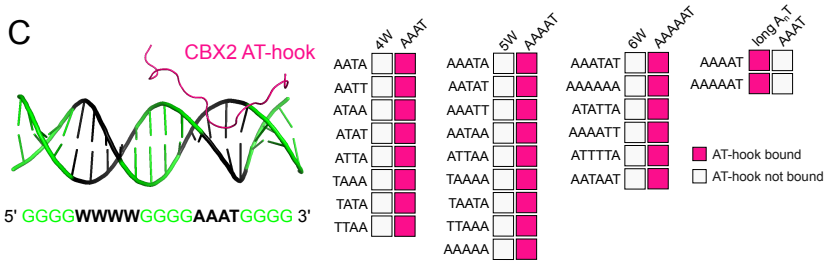
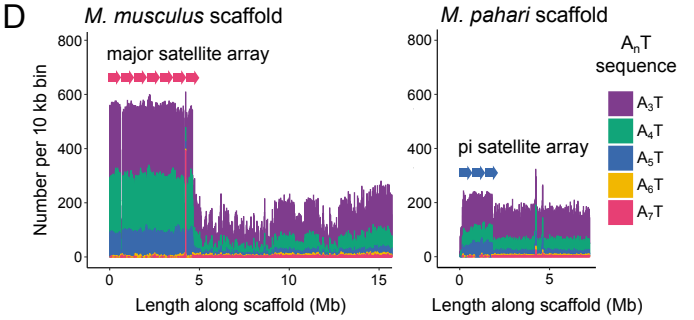

A
M. musculus ♂
M. pahari ♂
1
2
3
DAPI
H3K9me3
Survivin (CPC)
M. m ♂
M. p ♂
B
***
DAPI (Mean)
Distance along chromosome (microns)
M. musculus
M. caroli
M. pahari
C
Sytox Green
H3K27me3
Sytox
D
Sytox Green (Mean)

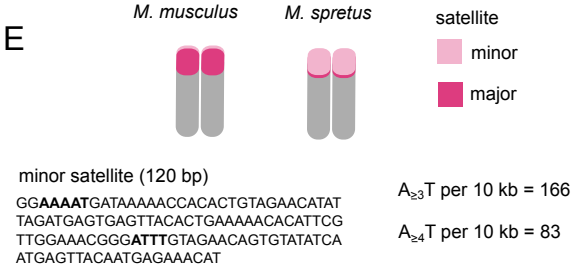
Figure 3

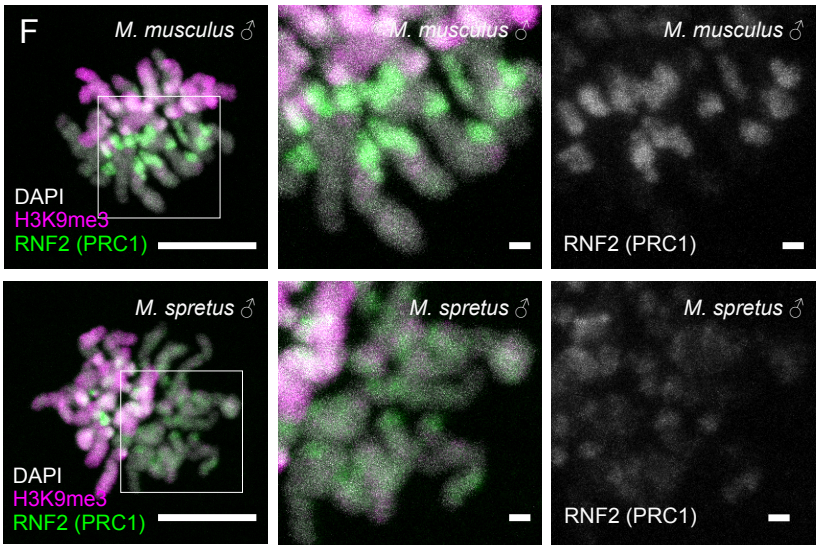

A
M. musculus scaffold
major satellite array
M. pahari scaffold
pi satellite array
A/T run length
4
5
6
7
8
9
≥10
Number per 10 kb bin
Length along scaffold (Mb)


B
4W sequence
AAAA
AAAT
AATA
AATT
ATAA
ATAT
ATTA
TAAA
TATA
TTAA
5W sequence
AAAAA
AAAAT
AAATA
AAATT
AAATAA
AATAT
AATTA
ATAAA
ATAAT
ATATA
ATTAA
ATTTA
TAAAA
TAATA
TATAA
TTAAA
Avg number per 10 kb bin
major
pi
Satellite array


C
CBX2 AT-hook
5' GGGGWWWWGGGGAAATGGGG 3'
4W
AAAT
AATA
AATT
ATAA
ATAT
ATTA
TAAA
TATA
TTAA
5W
AAAAT
AAATA
AAATT
AATAA
ATTAA
TAAAA
TAATA
TTAAA
AAAAA
6W
AAAAAT
AAATAT
AAAAAA
ATATTA
AAAATT
ATTTTA
AATAAT
long AnT
AAAAT
AAAAAT
AT-hook bound
AT-hook not bound


D
M. musculus scaffold
major satellite array
M. pahari scaffold
pi satellite array
AnT sequence
A3T
A4T
A5T
A6T
A7T
Number per 10 kb bin
Length along scaffold (Mb)


E
M. musculus
M. spretus
satellite
minor
major
minor satellite (120 bp)
GGAAAATGATAAAAACCACACTGTAGAACATAT
TAGATGAGTGAGTTACACTGAAAAACACATTCG
TTGGAAACGGGATTTGTAGAACAGTGTATATCA
ATGAGTTACAATGAGAAACAT
A≥3T per 10 kb = 166
A≥4T per 10 kb = 83


F
M. musculus ♂
DAPI
H3K9me3
RNF2 (PRC1)
M. spretus ♂

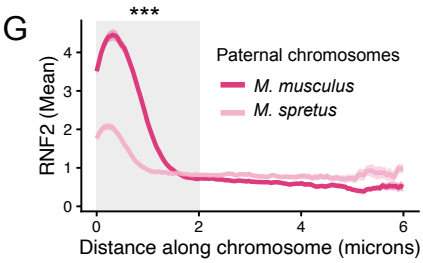
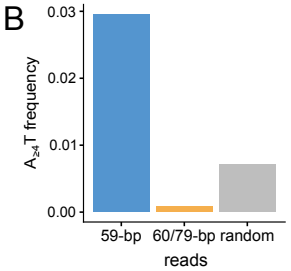
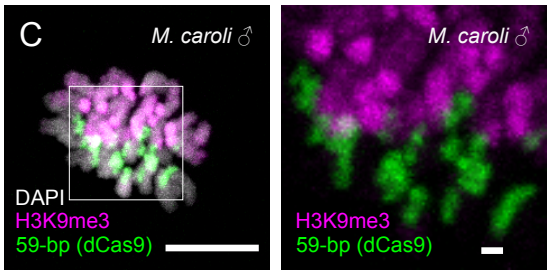

G
***
Paternal chromosomes
M. musculus
M. spretus
RNF2 (Mean)
Distance along chromosome (microns)

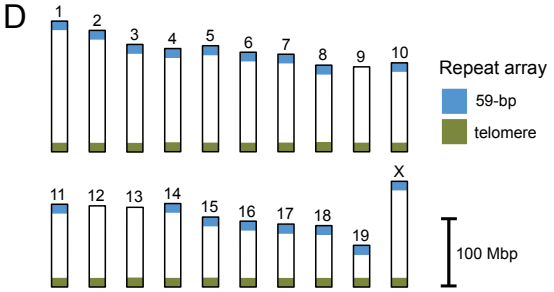
Figure 4

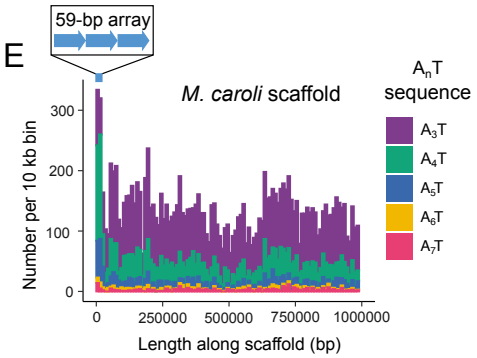
A

Cluster	Satellite Probability	% input reads	TAREAN consensus	Length (bp)	Theoretical A_nT per 10 kb array
1	0.979	5.3	CACAGTGTAGTACATTGTATA TGAGTGAGTTGCACTGAAA AACTTAGA**ATTT**ACTGAAAA ACCCGCATTAGACG**AAAT**G	79	253
2	0.992	6.0	CACACTGTAGGACCCGGAAT ATGGCGAGAAAACCTGAAAA AGGTGG**AAAAT**GAG**AAAAT**	59	338









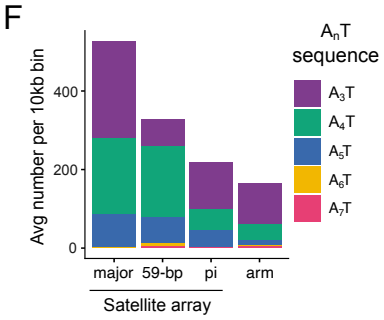
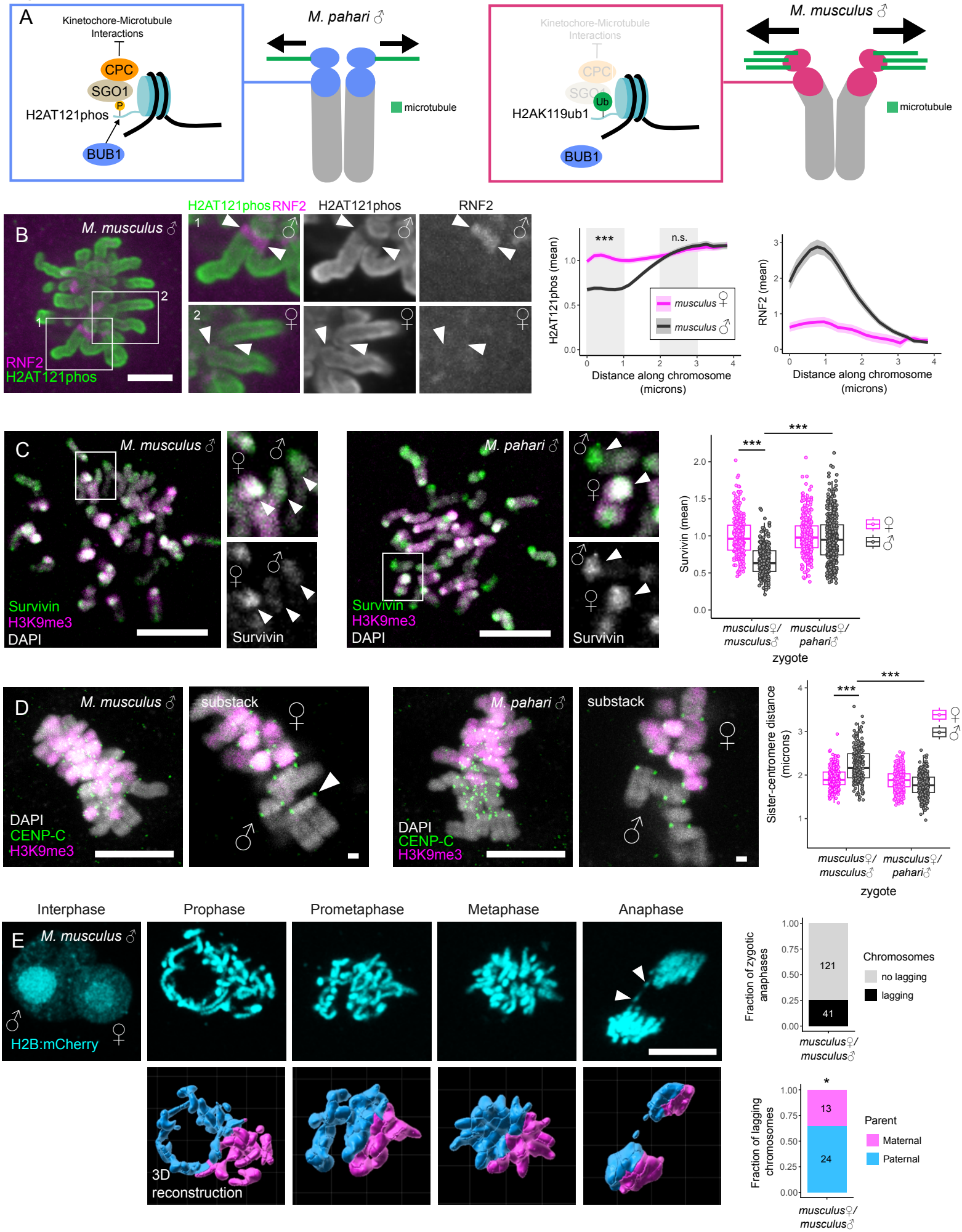


Figure 5


A
Kinetochore-Microtubule Interactions
CPC
SGO1
H2AT121phos
P
BUB1
M. pahari ♂
microtubule
Kinetochore-Microtubule Interactions
CPC
SGO1
Ub
H2AK119ub1
BUB1
M. musculus ♂
microtubule
B
M. musculus ♂
RNF2
H2AT121phos
H2AT121phos RNF2
H2AT121phos
RNF2
***
n.s.
H2AT121phos (mean)
musculus ♀
musculus ♂
Distance along chromosome (microns)
RNF2 (mean)
Distance along chromosome (microns)
C
M. musculus ♂
Survivin
H3K9me3
DAPI
Survivin
M. pahari ♂
Survivin
H3K9me3
DAPI
Survivin
***
***
Survivin (mean)
musculus♀/ musculus♂
musculus♀/ pahari♂
zygote
D
M. musculus ♂
substack
DAPI
CENP-C
H3K9me3
M. pahari ♂
substack
DAPI
CENP-C
H3K9me3
***
***
Sister-centromere distance (microns)
musculus♀/ musculus♂
musculus♀/ pahari♂
zygote
Interphase
Prophase
Prometaphase
Metaphase
Anaphase
E
M. musculus ♂
H2B:mCherry
3D reconstruction
Fraction of zygotic anaphases
121
41
Chromosomes
no lagging
lagging
musculus♀/ musculus♂
*
Fraction of lagging chromosomes
13
24
Parent
Maternal
Paternal
musculus♀/ musculus♂