Science Advances
AAAS

Supplementary Materials for

Centromere innovations within a mouse species

Craig W. Gambogi *et al.*

Corresponding author: Ben E. Black, blackbe@pennmedicine.upenn.edu

Sci. Adv. **9**, eadi5764 (2023)
DOI: 10.1126/sciadv.adi5764

This PDF file includes:

Figure S1

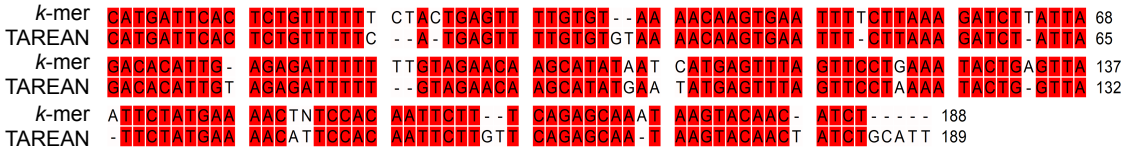
k-mer CATGATTCAC TCTCTTTTTT CTACTCAGTT TTGTGT--AA AACAAGTGAA TTTTCTTAAA GATCTTATTA 68
TAREAN CATGATTCAC TCTGTTTTTC --A-TGAGTT TTGTGTGTAA AACAAGTGAA TTT-CTTAAA GATCT-ATTA 65
k-mer GACACATTG- AGAGATTTTT TTGTAGAACA AGCATATAAT CATGAGTTTA GTTCCTGAAA TACTGAGTTA 137
TAREAN GACACATTGT AGAGATTTTT --GTAGAACA AGCATATGAA TATGAGTTTA GTTCCTAAAA TACTG-GTTA 132
k-mer ATTCTATGAA AACTNTCCAC AATTCTT--T CAGAGCAAAT AAGTACAAC- ATCT----- 188
TAREAN -TTCTATGAA AACATTCCAC AATTCTTGTT CAGAGCAA-T AAGTACAACT ATCTGCATT 189

Figure S1: π-sat sequence is almost identical to the top hit identified by the *k*-mer strategy.
Alignment of the satellites derived from the *k*-mer and TAREAN approach.

Figure S2

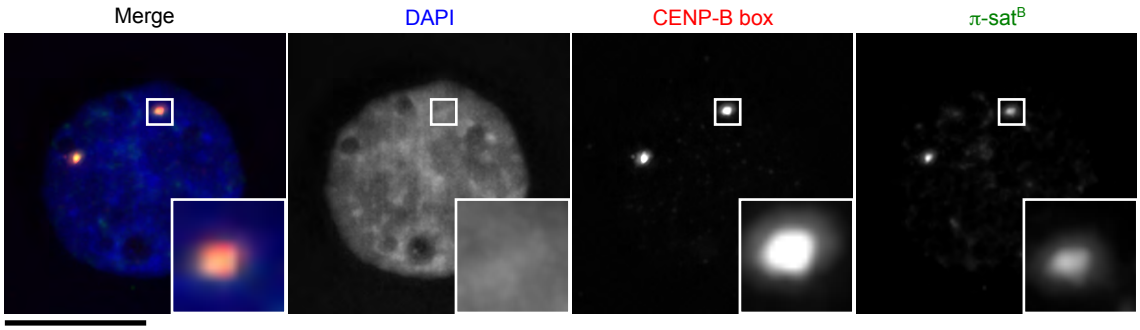

Figure S2: Functional CENP-B box is found at an pair of homologues containing π-sat^{B}.

Representative image of *M. pahari* fibroblast cells labeled with CENP-B box and π-sat^{B} FISH probes. Insets: 4.6x magnification. Bar, 10 μm.

Figure S3

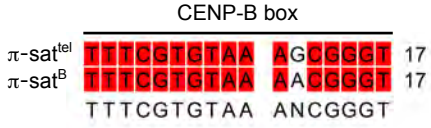
CENP-B box
π-sat^tel TTTCGTGTAA AGCGCGGT 17
π-sat^B TTTCGTGTAA AACGGGT 17
TTTCGTGTAA ANCGGGT

Figure S3: Functional CENP-B boxes found at chromosome 11 differ from functional CENP-B boxes found on π-sat^{tel} on other centromeres.

Alignment of functional CENP-B box from π-sat^{B} and π-sat^{tel}

Figure S4

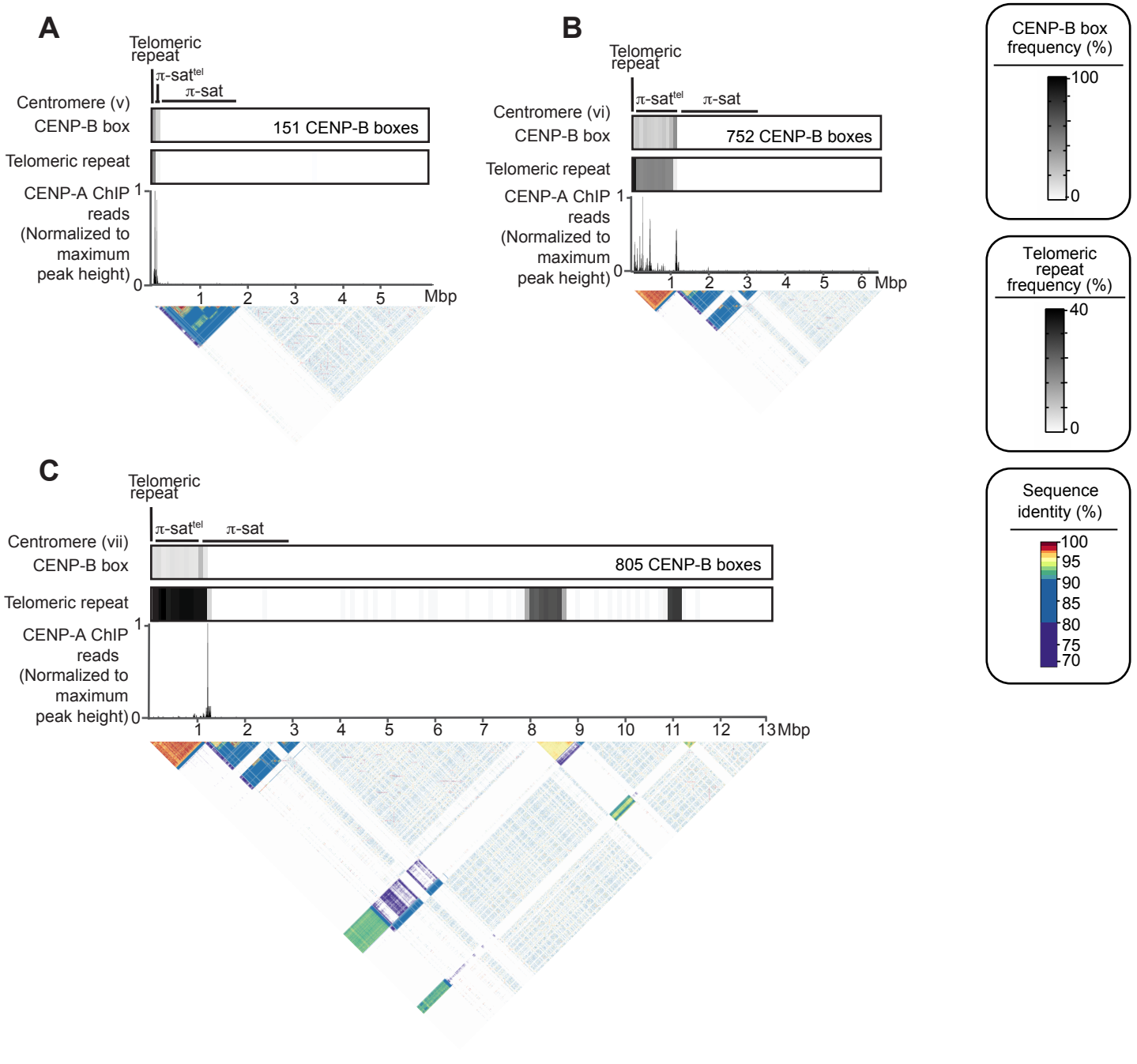
A
Telomeric repeat
π-sat^tel
π-sat
Centromere (v)
CENP-B box
151 CENP-B boxes
Telomeric repeat
CENP-A ChIP reads (Normalized to maximum peak height)
1
0
1 2 3 4 5 Mbp
B
Telomeric repeat
π-sat^tel
π-sat
Centromere (vi)
CENP-B box
752 CENP-B boxes
Telomeric repeat
CENP-A ChIP reads (Normalized to maximum peak height)
1
0
1 2 3 4 5 6 Mbp
C
Telomeric repeat
π-sat^tel
π-sat
Centromere (vii)
CENP-B box
805 CENP-B boxes
Telomeric repeat
CENP-A ChIP reads (Normalized to maximum peak height)
1
0
1 2 3 4 5 6 7 8 9 10 11 12 13 Mbp
CENP-B box frequency (%)
100
0
Telomeric repeat frequency (%)
40
0
Sequence identity (%)
100
95
90
85
80
75
70

Figure S4: Three additional *M. pahari* centromeres, all containing similar overall organization. A-C) The fraction of π-sat repeats containing a functional CENP-B box (NTTCGNNNNANNCGGGN) and the frequency of telomeric repeats (TTAGGG) are shown. CENP-A ChIP-seq reads were aligned to the assembly revealing that CENP-A is primarily present on π-sat^{tel}. A pairwise sequence identity heat map indicates the degree of homogeneity in centromeric DNA.

Table S1

Species	Repeat associated with CENP-A nucleosomes	Repeat unit size (bp)	Sequence identity to minor sat (%)
M. musculus	Minor sat	120	-
M. spretus	Minor sat	120	-
M. caroli	*M. caroli* satellite repeat	60 or 79	57
M. pahari chromosome 11	π-sat^{B}	189	52
Other *M. pahari* chromosomes	π-sat^{tel}	n/a	42

Table S1: Summary of centromere repeats in *M. musculus*, *M. spretus*, *M. caroli*, and *M. pahari*.

Various *Mus* species have major sequence differences at the centromere in terms of both sequence identity and repeat size. For other M. pahari chromosomes, π-sat^{tel} has an complex repeat (Fig. 4B), so it is listed as n/a.