- 1 CENP-B box, a nucleotide motif involved in centromere formation, occurs in a
- 2 New World monkey
- $\frac{3}{4}$

5 Electronic supplementary material

6

7 Supplementary methods

- 8 (Detailed description of the methods used in this study)
- 9
- 10 Immunofluorescent cell staining: Marmoset epithelial cells and human HeLa
- 11 cells were cultured in AminoMAX-II Complete Medium (Life technologies) and
- 12 D-MEM High Glucose (Wako) supplemented with 10% FBS and
- 13 penicillin-streptomycin, respectively, in a humidified incubator set at 37 °C and
- 14 5% CO₂. Immunofluorescence staining was performed as described previously
- 15 [14] with a few modifications. Cells grown in glass-bottom plates were fixed in
- 16 2.6% formaldehyde in PBS for 10 min at room temperature, and quenched with
- 17 0.4% Triton X-100 and 100 mM glycine in PBS for 7 min at room temperature.
- 18 The fixed cells were incubated in blocking solution, 1% BSA in PBS-T (PBS
- 19 containing 0.1% Triton X-100), for 30 min at room temperature. Primary
- 20 antibody incubation was performed in 0.1% BSA in PBS-T for 1 h at room
- 21 temperature and the cells were washed with PBS-T four times. Incubation with
- 22 fluorescently labeled secondary antibody was performed in 0.1% BSA in PBS-T
- for 1 hour at room temperature and the cells were twice washed with PBS-T.
- The cells were then incubated in PBS-T containing DAPI for 5 min at room
- 25 temperature. After washing with PBS-T and PBS, images were acquired using an
- LSM700 microscope (Zeiss) equipped with Alpha Plan-Apochromat 63×/1.46
- Oil M27 lens (Zeiss). Three slices of Z-stacks with a spacing of $0.34 \,\mu$ m were
- displayed as maximum intensity projections using Zen 2009 (Zeiss). Signals of
- 29 small sizes compared to the survey spacing might escape fluorescence detection,
- 30 possibly resulting in an underestimation of the signal number. The following
- antibodies were used. For CENP-A staining, rat anti-human CENP-A antibody
- 32 (6F2 [14]) and goat anti-rat IgG (H+L) secondary antibody, Alexa Fluor 694
- 33 conjugate (Life technologies). For CENP-B staining, mouse monoclonal
- anti-human CENP-B antibody (5E6C1) and goat anti-mouse IgG (H+L)

- 35 secondary antibody, Alexa Fluor 488 conjugate (Life technologies). Our
- 36 anti-human CENP-A antibody and anti-human CENP-B antibody can be thought
- to recognize marmoset CENP-A and CENP-B, respectively, because the amino
- acid sequences of marmoset CENP-A and CENP-B exhibit high similarities to
- their human equivalents: 91.4% and 96.6% identity, respectively [2].
- 40
- 41 Sequencing of fosmid clones for long contig sequences: The strategy of this
- 42 method was fully described in our previous reports [13, 18], and its essence is
- 43 described here. The fosmid clone to be sequenced (FosMar08) contained a 40-kb
- 44 AS fragment. At the boundary between the vector portion and the insert
- 45 fragment, there was a recognition site for the restriction endonuclease *Bam*HI.
- 46 The insert fragment did not contain a *Bam*HI site. We delivered a Tn5
- 47 transposon into this fosmid clone by inducing its transposition, and collected
- 48 several secondary clones. The Tn5 transposon had been modified in advance to
- 49 carry a *Bam*HI site. This modification enabled us to determine the position of
- 50 the Tn5 insertion points by digesting the secondary clones with *Bam*HI and
- 51 measuring the fragment size by gel electrophoresis. From the collection of the
- 52 secondary clones for which the insertion points were determined, we selected a
- total of 24 clones so that neighboring clones would be located within 1.0 kb. We
- 54 then sequenced the selected secondary clones by using a pair of primers that
- 55 corresponded to the internal region of the Tn5 transposon and were oriented
- 56 outwards, which resulted in sequence reads of more than 1.5 kb in length. By
- 57 assembling these sequence reads in the order of their positions, we obtained a
- 58 contig sequence of 13.1 kb in length.



- **Figure S1**. Phylogenetic structure of simian primates. The upper branch of each
- 62 taxonomic level is further divided into lower-level taxa (not shown).



- Figure S2. Structure of AS repeat units contained in the FosMar08 clone. The 66
- higher-order structure was found in our previous study [12], and sequence 67
- identities observed in this clone are shown. It is clear that repetition of a specific 68
- distribution pattern of CENP-B boxes coincides with the higher-order repeat 69
- structure of AS. 70



- $71\\72$
- **Figure S3.** Distribution of CENP-B boxes among repeat units contained in
- sequence reads. The names of the 24 fosmid clones are listed on the left of the
- 75 alignment. Partial or intact repeat units contained in their sequence reads are
- 76 schematically shown. The circles located near the right ends of the repeat units
- indicate sites for CENP-B box. Magenta-coloured and white circles indicate the
- 78 presence and absence of CENP-B boxes, respectively.

A-TTGAGAAA-TATCTGCTTCTAAAAC-CAAAACGGAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-AGAGTTAAA-CTGATG Con Mar102 A-TTGAGAAA-TAT-TGCTTCTAAACC---AAACGGAGCT-AC-TAACAGAACGGT-CT-CAATGT-GTGCATTCAA-CT-AC-GGAGTTTAA-TT-ATG Mar104 A-TTGAGCAA-TATCTGCTTCTAAAAAC-CAAAACGGAGCT-ACCTAACAGAACGGTCATCAA-TGT-GTGCGTTCAA-CGTAG-AGAGTTAAG-CTGATG Mar107 A-TTGAGAAA-TATCTGCTTCTAAAAC-CAAAACGGAGCA-AGCTAACAGAATGGTCATCAAATGT-GTCAGTTCAA-CTTAG-AGAGTTAAG-CTGATG Mar108 A-TTGAGAAA-TATCTGCTTCTAGAAC-AAAAACGCAGCT-ACGTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-AGAGTTAAA-CTGATG Mar112 T-TGGAGAAA-TATCTGCTTCTAAAAAC-CAAAACGGAGCA-ACCTAACAGAACGGC-CTTCAATGT-GTGCGTTCAA-CTTAC-AGAGTTAAA-TTGATG Mar113 A-TTGAGAAAATATCTGCTTCTAGAACCAAAAACGCAGGT-TCGTTACAGAACGGT-CTTCAAAGT-GTGCATTCAA-CTTAC-AGAGTTAAA-AGGATG Mar01A7 Mar01A8 A-TTGAGAAA-TATCTGCTTCTAAAAC-CAAAACGGAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-GGAGTTAAA-CTGATG Mar01E4 A-TTGAGAAAATATCTGCTTCTAAAAC-CAAAACGGAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-AGAGATAAA-CTGATG Mar01E7 Mar01G2 A-TTGAGAAA-TATTT-CTTTTGAAACCCAAAAACGGAGCT-ACCTAACAGAGTGGTCTTCAAACGT-GTGCATTCAA-CTTACAGAAGTTAAAAATAATTG Mar01K7 A-TTGAGAAA-TATCTGCTTCTAGAAC-CAAAACGCAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-AGAGTTAAA-CTGATG Mar02B3 A-TTGAGAAA-TATCTGCTTCTAAAAC-CAGAACGGAGCA-ACTTAACAGAACGGC-CTTCAATGT-GTGCGTTCTA-CTTAC-AGAGTTGAA-CTGATG Mar02L2 A-TTGAGAAA-TATCTGCTTCTAAAAC-CAAAACGGAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-GGAGTTAAA-ATGATG Mar32E3 A-TTGAGAAA-TATCTGCTTCTAAAAC-CAAAACGGAGCT-ACCTAGCAGAACGCTCTTCAA-TGT-GTGCGTTCAA-CTTAC-AGAGTTAAG-CTGGTG Mar32J3 Mar 33H3 A-TTGAGAAA-TATCTGCTTCTAGACC-CAAAACGCAGCT-A-CTAAGAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-AGAGTTAAAACTGATG A-TTGAGAAA-TATCTGCTTCTAAAAC-CAAAACGGAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-GGAGTTAAA-CTGATG Mar34I2 Mar34D8 Mar35H2 A-TTGAGAAA-TATCTGCTTCTAAAAC-CAAAACGGAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-GGAGTTCAA-CTGATG Mar35J8 a-ttgagaaa-tatctgcttctaaaac-caaaacggagct-acctaacagaacggt-cttcaatgt-gtgcattcaa-aaaag-ggagttaaa-ctgatg Mar36C3 Mar36D4 T-TGGAGAAA-TATCGGCTTCTAAAAC-CAGAACGGAGCA-ACCTAACAGAACGGC-CTTCAATGT-GTGCGTTCAA-CTTAC-AGAGGTAAA-CTGATG Mar02D4 A-TTGAGAAA-TATCTGCTTCTAAAAA-CAAAACGGAGCT-ACCTAACAGAACGGT-CTTCAATGT-GTGCATTCAA-CTTAC-GGAGTTAAA-CTGATG TGT-GTTT-GCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGCG-GAAAACGCA-TTTTCCCAG--CCAAATCC-AAGCATTTAAG--AAATAGCTG TGT-GTTC-GCAGGAGTTT-CTAT-CCC--TTTCTGTCAA-GAAGTG-GAAA-CGCA-TTTTCCCAG--CCAAATCC-AGGCATTCAAG--AAATAGCTG TGT-GTTT-GCAGGAGTTT-GTAAACCCTTT-CTGTGAA-GAAGGG-GAAAAAGCA-TTTT-CCCCGG-CGAAATCC-AAGCATTTAAG-AAATAGCTG TGT-GTTTTGCAGGAGTTT-GTAAACCCCTTT--CTGTGAA-GAATCG-GAAAAAGCA-TTTT-CCCGG-CCAAATCC-AAGCATTTAAA--AAATAGCTG TGC-GTTT-GCAGGAGTTT-GTAAACCC--TTTCTGTGAA-GCAGCGGAAAA-CGCA-TTTTTGAAG--CCAAATCC-AAGCATTTCAG--AAATAGCTG TGT-GTTT-GCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGCG-CAAAACGCA-TTTTTCCAG--GCAAAAGC-CAGCATTTCCCC--AAATAGCTG TGC-GTTT-GCAGGAGTTT-GTAAACCC--TTTGTGTGGAA-GCAGCGGAAAA-CGCA-TTTTTCCAG--CCAAATCC-AAGCATGTCAG--GAATAGCTG TGT-GTTT-GCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGCG-CAAAACGCA-TTTTCCCAG--GCAAAAGC-AAGCATTTCCC--AAATAGCTG TGT-GTTT-GCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGTG-GAAAACGCA-TTTTCCCAG--CCAAATCC-AGGCATTCAAG--AAATAGCTG TGT-GTTT-GCAAGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGTG-GAAAACGCC-TTTTCCCAG--CCAAATGC-AGGCATTCAAG--AAATAGCTG TGT-GTT--GCAGTTAGTTTCTAGACGC-TTTTTTGTGAA-GATGCA-GAAAACGCG-TTTTTACAG--CCAAATCC-AAGTATTTCAG--CAATAGCGG TGC-GTTT-GCAGGAGTTT-GTAAACCC--TTTCTGTGAA-GCAGCGGAAAA-CGCA-TTTTTGCAG--CCAAATAC-AAGCATTTCAG--AAATAGCTG TGT-GTTT-GCAGGAGTTT-GCAAACTT-TTTCTGTGAA-GAAGCA-GAAAACGAA-TGTTTCCAG--CCAAATCC-AAGCATTTCAG--AAATAGCTG TGT-GTTT-GCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGTG-GAAAACGCA-TTTTCCCAG--CCAAATCC-AGGCATTCAAG--AAATAGCTG TGC-GTTT-GCAGGAGTTTTGTAAACCC--TTTCTGTGAA-GCAGCGGAAAAACGCAATTTTTGCAG--TCAAATAC-AAGCATTTCAA--AAATATCTG TGTCGTTTTGCAGGAGTTT-TTAAACCCCTTTTTTGTGAA-GAAGTG-GAAAACGCA-TTTTCCCCAGCCCAAATGC-AGGCATTCAAG--AAATAGCAG TGT-GTTT-GCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GGAGTG-GAAAACGCT-TTTTCCTAG--CCAAATGC-AGGCATTCAAT--AAATAGCTG TGT-GTTTTGCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGTG-GAAAACGCC-TTTTCCCAG--CCAAATCC-AGGCATTCAAG--AAATAGCTG TGT-GTTC-GCAGGAGTTT-CTAAATTC--TTTCTGTGGA-GAAGTG-GAAAACGCA-TTTTCCCAG--CCAAATGC-AGGCATTCAAG--AAATAGCTG TGT-GTTT-GCAGGAGTTT-CTAAACAT-TTTCTGTGAAAGAAGCA-AAATACGCT-ATTTCCCAG--CAAAATGC-AAGCATTTCCCC--AAATAGCTG TGT-GTTT-GCAGGAGTTT-CTAAACCC--TTTCTGTGAA-GAAGTG-GAAAACGCA-TTTTCCCAG--CCAAATCC-AGGCATTCAAG--AAATAGCTG TAAA-TCTCCGTCTGAAACACAAAACGAA-CGTATCCTTCAGA-CTGCTCGCTGGGTGTTCGCTGACCTAAGGGAGTTGAAT-CTGCCTCTAGAT-TCA TAAA-TCTCCCGTCTAAAAACACAAAACGAA-CGTATCCTT-CAAACTGATCTGCTGTGCGTTCGCTTAAATAAGAGAGTTGAAT-CCGGGTTTAGAT-TCA TAAA-TCTCCGTCTGAAACACAAAACGAA-CGTATCCTTCAGA-CTGCTCTCGGGTGGTGGCTTAGCTAAGGGAGTTGAAT-CTGCCTCTAGAT-TCA GAAA-TCTCC---TAAAACACAAAAAAGAA-CATATCTTGCAGA-CTGCTCGCGGGGGTTGGCTCATCTAATGAAGTAGAAT-CCGCCTTCTGAA-GAA GAAA-TCTCCGTCTAAAAACACAAAAAAGAA-CGTATCTTGAAAA-CTGCTCGGGTGTTCGGCTCAACTAAGAAAGTTGAAT-CCGCCTTCAGAA-TCA GAAA-TCTCCCGTCTAAAACACAAAAAAGAA-CGTATCTTGCAAA-CTGCTCGGTGGGGGGTGTCGGCTAAGGGAGTTGAAT-CCGCCTTCAGAA-TCA TAAAATCTCCGTCTAAAACACCAAACAAAACGTATCTTCCAAA-CTGCTCTACTGTGTGTTGGTTCAACAAACGGAGTTGAAAGCTTTGTTTAGAT-TCA TAAA-TCTCCGTCTAAAAACACAAAAACGAA-CGAATCCTT-CAAACTGCTCTCCTGTGCGTTCGCTTAACTAAGGGAGTTGAAT-CCCCCGTTTAGAT-TCA GAAA-TCTCCGTCTAAAAACACAAAAAAGAA-CGTATCTTGCAAA-CTGCTCGGGTGTTCGGCTCAGCTAAGGGAGTAGAAT-CCGCATTCAGAA-TCA GAAA-TCTCCCGTCTAAAAACACAAAAACGAA-CGTTTCCTTCTAA-CTGCTCGCGTGCTTCGCCTAACTAAGGGAGTTGAAT-CCGCGTTTAGAC-GCA GAAA-TCTCCGTCTAAAAAAAAAAAGGA-CGTATCTTGCAAA-CTGCTCGGTGGGTGTTCGCTCAGCTAAGGGAGTTGAAT-CCGCCTTCAGAA-TCA GAAA-TCTCCCGTCTAAAAACACAAAAAAGAA-CGTATCTTGCAAAAACTGCTCGGCTGGGTGTTCGCTCAGCTAAGGGAGTTGAAT-CCGCCTTCAGAA-TCA GAAA-TCTCCGTCTAAAACACAAAAAGAA-CGTATCTTGCAAA-CTGCTCTGCAGGGTGTTCGCTCAGCTAAGGGAGTTGAAT-CCGCCTTCAGAA-TCA TAAA-TCTACGTCTAAAACACAAAACGAA-CGTATCCTTCAGA-CTGCTCGCTGGGTGTTCGCTTAACTAAGGGGGTTGAAT-CTGCCTCTAGAT-TCA GAAA-TCTCCGTCTAAAACACAAAAAGAA-CGTATCTTGCAAA-CTGCTCGCTGGCTGTTCGCTCAGCTAAGGGAGTTGAAT-CCGCCTTCAGAA-TCA

79 80

81 Figure S4. Alignment of entire sequences of first intact repeat units. The names

82 of the fosmid clones are listed on the left of the alignment. For additional

information see the legend to Figure 1.

+++++++
GAGAGTGAGAAGCACGTTTCTTCCAGGAGCTGCGTTTTGTTATACC?GGGTGAAT-GGGAGTAAACGCTGT
GAGAGTGAGAAGCACGGTTTTTCCAGGAGCTGGGTTTCGTTATACCCGGGTGCAT-TGGAACAAACGCGGT
GAGAGTGAAAAGCA-GTTTCTTCCAGGAGCTGCGTTTTGGTATTCCAGAGTGAAT-GGGAGTAAACGCTGT
TAGAGTGAGAAGCA-GTTTCTTCCAGGAGCTGCGTTTTGGTATTCCAGAGTGAAT-GGGAGTAAACGCTGT
GAGAGTGGGAAGCACGTTCCTTCTAGGAGCTGCGTTTTGCTATTCCAGCGTGAAT-GGGAGTAAACGCTGT
GAGAGTGAGAAGCACGTTGCTTCTCGAAACTGCGTTTTGATATCCCAGAGTGAAT-GGGAGTAAACGCGGT
GATAGTGGGAAGCACGTTCCTTCTAGGAGCTGCGTTTTGCTATTCCAGCGTGAAT-GGGAGTAAACGCTGT
GAGAGTGAGAAGCACGTTGCTTCTCGCAGCTGCGTTCTGATATGCCAGAGTGAAT-GGGAGTAAACGCGGGT
TAGAATGAGAAGCACGGCTCTTCCTCGAGCTGGATGTCGTAATACCCGGGGGAAT-TGGAGCAAACGCCGT
GAGAGTGAGAAGCACGTTTCTTCCAGGAGCTGCGTTTCCTTATACCCGGGGTGAAT-GGGAGCAAACGCGGT
GAGAGTGAGAAGCACGGTTCTTCCAGGAGATGTGTTTCGTTATACCCGGGTGAAT-GGGTGCAAACGCGAT
GAGAGTTTGAAGCACGTTTCTCCTTTGAGCTGCTTTTGTATATTCCCGAGCGAAT-GGCAGTAAACGCTGT
GAGAGTGGGAAGCACGTTCCTTCTAGGAGCTGCGTTTTCCTATTCCAGCGTGAAT-GGGAATAAACGCTGT
GAGAGTGGGAAGCACGTTCCTTCTAGGAGCTGCGTTTTGCTATTCCAGCGTGAAT-GGGAGTAAACGCTGT
GAGGGTGAGAAGCGCGTTTTTTCTAGGAGCTGCGTTTTGATATTCCAGAGAAT-GGGAGTAAACGCTGT
GAGAGTGGGAAGCACGGTCCTTCCAGGAGCTGGGTTTCGTAATACCCGGGTGAAT-TGGAACAAACGCAGT
GAGAGTGAGAAGCA-GTTTCTTCCAGGAGCTGCGTTTTGGTATTCCAGAGTGAAT-GGGAGTAAACGCTGT
GAGAGTGGGAAGCACGTTCCTTCTAGGAGCTGCGTTTTGCTATTCCAGCGTGAAT-GGGAGTAAACGCTGT
GAGAGTGAGAAGCACGTTTCGTCCAGGAGCTGCGTTTCGTTATACCCGGTTGAAA-TGGAACAAACGCTGT
GAGAGTGAGAAGCACGGTTCTTCCAGGAGCTGGGTTTCGTTATACCCGGGTGAATTGGGAACAAACGCATT
GAGAGTGAGAAGCACGGTTCTTCCAGGAGCTGTGTTTCGTTATACTGCGGTGAAT-GGGTGCAAACGCGAT
GAGAGTGAGAAGCACGGTTCTTCCAGGAGCTGCGTTTCGTTATACCCGGGTGATT-GGGTACAATCCCAAT
GAGAGTGAGAAGCACGGTTCTTCCAGGAGCTGTGTTTCGTTATACCCGGGTGAAT-GGGTGCAAACGCGAT
GAAAGTAAGAAGTCCGTTGCTTCTTGGAGCTGCGTTTTGACATGCCACACTGAAA-GGTAGTAAACACGGT
GAGAGTGAGAAGCACGGTTCTTCCAGGAGCTGGGTTTCGTTATACCCGGGTGAAT-TGGAACAACGCGGT

86 Figure S4 (continued)