

Birth, evolution, and transmission of satellite-free mammalian centromeric domains

Solomon G. Nergadze,^{1,6} Francesca M. Piras,^{1,6} Riccardo Gamba,^{1,6} Marco Corbo,^{1,6} Federico Cerutti,^{1,†} Joseph G.W. McCarter,² Eleonora Cappelletti,¹ Francesco Gozzo,¹ Rebecca M. Harman,³ Douglas F. Antczak,³ Donald Miller,³ Maren Scharfe,⁴ Giulio Pavesi,⁵ Elena Raimondi,¹ Kevin F. Sullivan,² and Elena Giulotto¹

¹Department of Biology and Biotechnology "Lazzaro Spallanzani," University of Pavia, 27100 Pavia, Italy; ²Centre for Chromosome Biology, School of Natural Sciences, National University of Ireland, Galway, H91 TK33, Ireland; ³Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University, Ithaca, New York 14850, USA; ⁴Genomanalytik (GMAK), Helmholtz Centre for Infection Research (HZI), 38124 Braunschweig, Germany; ⁵Department of Biosciences, University of Milano, 20122 Milano, Italy

Mammalian centromeres are associated with highly repetitive DNA (satellite DNA), which has so far hindered molecular analysis of this chromatin domain. Centromeres are epigenetically specified, and binding of the CENPA protein is their main determinant. In previous work, we described the first example of a natural satellite-free centromere on *Equus caballus* Chromosome II. Here, we investigated the satellite-free centromeres of *Equus asinus* by using ChIP-seq with anti-CENPA antibodies. We identified an extraordinarily high number of centromeres lacking satellite DNA (16 of 31). All of them lay in LINE- and AT-rich regions. A subset of these centromeres is associated with DNA amplification. The location of CENPA binding domains can vary in different individuals, giving rise to epialleles. The analysis of epiallele transmission in hybrids (three mules and one hinny) showed that centromeric domains are inherited as Mendelian traits, but their position can slide in one generation. Conversely, centromere location is stable during mitotic propagation of cultured cells. Our results demonstrate that the presence of more than half of centromeres void of satellite DNA is compatible with genome stability and species survival. The presence of amplified DNA at some centromeres suggests that these arrays may represent an intermediate stage toward satellite DNA formation during evolution. The fact that CENPA binding domains can move within relatively restricted regions (a few hundred kilobases) suggests that the centromeric function is physically limited by epigenetic boundaries.

[Supplemental material is available for this article.]

Chromosome segregation during mitosis and meiosis is directed by the centromere, the chromosomal locus that specifies kinetochore assembly during cell division (Cleveland et al. 2003; McKinley and Cheeseman 2015). Although the mechanism of kinetochore function in mitosis is highly conserved, centromere-associated DNA sequences are highly variable in evolution, a situation that has been referred to as the centromere paradox (Eichler 1999; Henikoff et al. 2001). In most multicellular organisms, centromeres are associated with large arrays of tandemly iterated satellite DNA sequences, typified by alpha-satellite DNA of primates in which a 171-bp sequence is present in arrays of up to megabase size at the primary constriction of mitotic chromosomes (Hayden et al. 2013). Despite this common theme, the sequences of the centromeric satellite DNA are divergent and are estimated to be among the most rapidly evolving components of the genome (Plohl et al. 2014). Direct evidence that DNA sequence is not the sole factor in determining centromere position or function was originally derived from examination of human chromosomal abnormalities. Dicentric chromosomes possessing kinetochore activity at only one of two alpha-satellite loci revealed that satellite

DNA is not sufficient for centromere specification (Earnshaw and Migeon 1985). Identification of anaphoid chromosomes, that nonetheless possessed fully functional centromeres, demonstrated that satellite DNA is not necessary for centromere function (Voullaire et al. 1993). Rather than DNA sequence, the common feature that links centromere function in most eukaryotes is the presence of a distinctive histone H3 variant, CENPA, which can directly confer centromere function to a locus when tethered experimentally (Palmer et al. 1991; Stoler et al. 1995; Mendiburo et al. 2011). These observations have led to the proposal that centromere identity is established and maintained through epigenetic mechanisms, and CENPA functions as a central component in centromere specification (Karpen and Allshire 1997; Panchenko and Black 2009; McKinley and Cheeseman 2015).

The evolutionary plasticity of centromeres is exemplified by the phenomenon of centromere repositioning (Montefalcone et al. 1999). By detailed molecular characterization of karyotypic relationships among primate species, it was observed that centromere position can change without a corresponding change in DNA organization (Montefalcone et al. 1999; Cardone et al. 2006; Ventura et al. 2007). In these cases, referred to as

⁶These authors contributed equally to this work.

[†]Deceased.

Corresponding authors: elena.giulotto@unipv.it, kevin.sullivan@nuigalway.ie, elena.raimondi@unipv.it

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.231159.117>.

© 2018 Nergadze et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.html>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

evolutionarily new centromeres (ENCs), centromere evolution seems to be driven by forces other than the surrounding DNA.

A relationship between ENCs and the anaphoid neocentromeres observed in human clinical samples emerged from analysis of the positions in which these events occur. For example, human neocentromeres at Chromosomes 3, 9, and 6 occur in the same genomic regions as ENCs observed in some primates, indicating that certain regions of the genome have a propensity to form centromeres (Ventura et al. 2004; Capozzi et al. 2008, 2009). Thus, regions of the genome may harbor “latent” centromere potential (Voullaire et al. 1993). The observation that the primate ENCs possessed typical arrays of alpha-satellite DNA led to the hypothesis that epigenetic marks can drive the movement of centromere function to new genomic sites, which can subsequently mature through the acquisition of satellite DNA sequences (Amor and Choo 2002; Piras et al. 2010; Kalitsis and Choo 2012). Following their original discovery in primates, a surprisingly large number of ENCs were identified in the genus *Equus* (Carbone et al. 2006; Piras et al. 2009), and some examples were also observed in other animals (Ferrerri et al. 2005; Kobayashi et al. 2008) and in plants (Han et al. 2009), indicating that centromere repositioning is a widespread force for karyotype evolution.

A fundamental step in understanding centromere biology was the discovery that the ENC at horse Chromosome 11 is completely devoid of satellite DNA (Wade et al. 2009). This observation revealed, for the first time, that a satellite-free centromere can be present in all individuals of a vertebrate species as a normal karyotype component. This centromere is established on a segment of DNA, conserved in vertebrates, which is free of genes as well as of satellite DNA, providing an example of an evolutionarily “young” ENC that has not acquired repetitive sequences. Satellite-free centromeres were subsequently observed in chicken (Shang et al. 2010), orangutan (Locke et al. 2011), and potato (Gong et al. 2012).

Examination of the centromere of horse Chromosome 11 in several individuals revealed that the satellite-free centromeric domains are present in each case, but the precise location of the CENPA binding region (~100 kb in length) differs among individuals and even between the two homologous chromosomes of a single individual (Purgato et al. 2015). Centromere activity could be associated with any sequence within a ~500-kb domain in the centromere forming region of Chromosome 11. Therefore, this “centromere sliding” is DNA sequence independent, as expected for an epigenetically defined locus. Thus, centromeres exhibit large-scale relocalization (centromere repositioning) during evolution as well as short-range relocalization (centromere sliding) within a population (Giulotto et al. 2017).

The genus *Equus* comprises eight extant species (two horses, three donkeys, and three zebras) that diverged from a common ancestor ~4 million years ago (Mya) (Steiner et al. 2012; Orlando et al. 2013). In a previous work, we analyzed the karyotype of four *Equus* species by in situ hybridization with satellite DNA probes and revealed that, in the domestic donkey (*E. asinus*) and in two zebras (*E. burchelli* and *E. grevyi*), a large number of centromeres lack detectable satellite DNA (Piras et al. 2010; Geigl et al. 2016), whereas in the horse, Chromosome 11 is the only one.

The aim of this work was to verify the presence of satellite-free centromeres in *E. asinus*, using ChIP-seq with anti-CENPA antibodies, to analyze their DNA sequence organization, positional stability, and transmission.

Results

Satellite-free CENPA binding domains in *Equus asinus*

Our previous work identified several donkey centromeres that lack detectable satellite repeats (Piras et al. 2010). Here, to identify the DNA sequences at these centromeres, ChIP-seq experiments were carried out on donkey primary skin fibroblasts. Two different antibodies were used to immunoprecipitate formaldehyde cross-linked chromatin fragments: a rabbit antiserum against CENPA (Wade et al. 2009) and a human CREST serum with high titer against CENPA (Purgato et al. 2015; Cerutti et al. 2016). DNA purified from immunoprecipitated and input chromatin was then subjected to paired-end Illumina sequencing. Since we previously demonstrated the presence of a satellite-free centromere on horse Chromosome 11 by ChIP-on-chip (Wade et al. 2009; Purgato et al. 2015), as positive control, we carried out the same ChIP-seq experiment with chromatin from horse skin fibroblasts. The horse and donkey genomes share an average of >98% sequence identity (Orlando et al. 2013; Huang et al. 2015) and chromosome orthologies are well described (Yang et al. 2004; Musilova et al. 2013). Since only draft sequences of the donkey genome comprising unassembled scaffolds are available (Orlando et al. 2013; Huang et al. 2015), we aligned both the horse and the donkey reads to the horse reference genome (EquCab2.0). Sequencing and alignment statistics of the ChIP-seq experiments are reported in Supplemental Table S1. Figure 1 reports the graphical representation of the enrichment peaks, corresponding to the centromere of horse Chromosome 11 from one individual, here called HorseS (Fig. 1A), and to the 16 donkey satellite-free centromeric domains from one individual, here called DonkeyA (Fig. 1B). The two antibodies recognized essentially identical sequence domains and exhibited largely similar patterns of protein binding.

The 16 donkey regions spanned 54–345 kb and contained one or two CENPA binding domains. Similar to what we described for horse Chromosome 11 (Purgato et al. 2015), the presence of two peaks is related to different epialleles on the two homologs, as demonstrated below on the basis of single nucleotide variant (SNV) analysis. Although some peaks showed a Gaussian-like regular shape (such as EAS4 and EAS30), other peaks were irregular (such as EAS8 and EAS14), contained gaps (such as EAS7 and EAS14), or exhibited a narrow, spike-like distribution (such as EAS9 and EAS19).

The satellite-based donkey centromeres are not described here because their corresponding ChIP-seq reads cannot be precisely mapped on specific chromosomes in the horse reference genome. These centromeres are probably organized similarly to the great majority of typical mammalian centromeres, as already shown for satellite-based horse centromeres (Nergadze et al. 2014; Cerutti et al. 2016).

CENPA binding domains correspond to primary constrictions in 16 *E. asinus* chromosomes

Cytogenetic analysis was carried out to map the 16 donkey CENPA binding regions relative to the primary constrictions of horse and donkey chromosomes. CENPA binding domain coordinates were used to select a set of horse BACs from the CHORI-241 library (Supplemental Table S2; Leeb et al. 2006). These were used as probes for in situ hybridization on metaphase spreads of horse and donkey skin fibroblasts. Examples of in situ hybridization

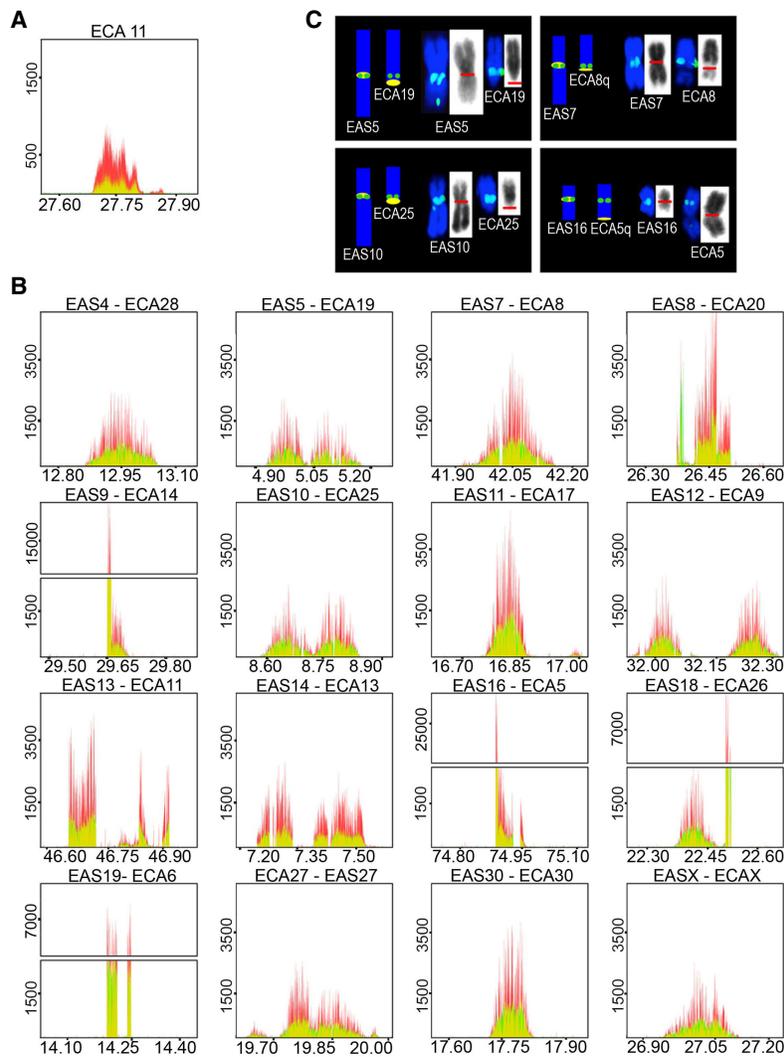


Figure 1. Identification of satellite-free centromeres in *Equus asinus*. ChIP-seq reads from primary fibroblasts of HorseS (A) and DonkeyA (B) were mapped on the EquCab2.0 horse reference genome. Immunoprecipitation was performed with an antibody against human CENPA (red) or with a CREST serum (green). Peak overlapping appears in yellow. The y-axis reports the normalized read counts, whereas the x-axis reports the genomic coordinates (Mb). The *E. caballus* satellite-free centromere from Chromosome 11 (A) and the 16 satellite-free *E. asinus* centromeres (B) are shown; for each *E. asinus* (EAS) chromosome, the number of the orthologous *E. caballus* chromosome (ECA) is reported. (C) FISH with BAC probes covering the genomic regions identified by ChIP-seq. Four examples (EAS) along with their orthologous horse chromosomes (ECA) are shown; the remaining chromosomes are reported in Supplemental Figure S1. On the left of each panel, a sketch of the orthology between *E. caballus* and *E. asinus* chromosomes (Yang et al. 2004; Musilova et al. 2013) is shown, with BAC signals represented as green dots, and the position of the cytogenetically determined primary constriction represented as a yellow oval. On the right of each panel, metaphase chromosomes are shown with FISH signals in green, and the primary constriction is marked by a red line on the reverse DAPI images (gray).

results are shown in Figure 1C with remaining data presented in Supplemental Figure S1. Each of the BAC probes identified a unique locus on the donkey karyotype, and its location was always consistent with the location of the primary constriction. Notably, the FISH signal on the orthologous horse chromosome was never centromeric, suggesting that the 16 satellite-free donkey centromeres were repositioned during evolution. We conclude that the 16 CENPA binding domains identified by ChIP-seq analysis are ENC located within the respective cytogenetically defined primary constrictions.

Sequence assembly of satellite-free centromere domains and comparison with orthologous horse genomic regions

Several CENPA binding domains showed read-free gaps and distorted shapes when mapped to the horse reference genome, suggesting differences in DNA sequence between the two species (Fig. 1B). The actual DNA sequence corresponding to the donkey centromeres was determined by assembling Illumina reads and carrying out Sanger sequencing of selected regions to resolve gaps in the assembly. For each centromeric region, genomic segments ranging in size between 157 and 358 kb were assembled (Supplemental Table S3).

In the majority of donkey satellite-free centromeres, multiple rearrangements (deletions, insertions, and inversions) were observed compared to the horse orthologous sequence (EAS4, EAS5, EAS7, EAS10, EAS11, EAS12, EAS13, EAS14, EAS27, EAS30) (Supplemental Fig. S2). The number and size of these rearrangements varies at different centromeres, but deletions are the most prevalent type. In donkey Chromosome 5, we observed several deletions; given the small size of these deletions, no gaps in the peak profile were observed. Conversely, donkey Chromosome 7 contains three relatively large deletions coinciding with gaps in the peak profile. The organization of the centromere of donkey Chromosome 13 is more complex, including a large deletion (110 kb) and a translocation, giving rise to a large gap in the central region (deletion) and an off-site peak outside the right border (translocation). In EAS14, which shows a two-peak profile, four relatively extended deletions coincide with gaps in the peak profile. No rearrangements were evident in the centromere of donkey Chromosome X. The centromeric domain identified by ChIP-seq is contained within the previously described large pericentric inversion of donkey Chromosome X (Raudsepp et al. 2002).

To determine more precisely the organization of CENPA distribution at satellite-free centromeres, we constructed a chimeric reference genome by inserting the assembled centromeric donkey contigs in EquCab2.0 to replace their orthologous horse sequences (Supplemental Table S3). The result was a virtual reference genome named EquCabAsiA.

ChIP-seq reads were then mapped on the EquCabAsiA genome (Supplemental Fig. S3). Comparison of the peak profiles obtained with the two reference genomes (Fig. 1B; Supplemental Fig. S3) shows that large gaps and irregular profiles that were observed in Figure 1B (EAS7, EAS13, EAS14, EAS16, EAS19) were no longer

detected following the new alignment. These results demonstrate that the CENPA binding domains of the satellite-free donkey centromeres are uninterrupted, and their architectural organization resembles that of horse Chromosome 11 (Fig. 1A; Wade et al. 2009).

Tandem repetitions associated with some satellite-free centromeres

For five donkey centromeres (EAS8, EAS9, EAS16, EAS18, and EAS19), we detected novel tandem repetitions of sequences that are single copy in the horse genome. In particular, reads spanning junctions between adjacent units of tandem arrays directly demonstrated their presence. For EAS18 and EAS19, the amplified sequences contain a deletion relative to the horse genomic sequence (Supplemental Fig. S2). Due to their repetitive nature, these five regions could not be precisely assembled. To prove the presence of tandem repetitions at these centromeres and to determine their copy number, three independent approaches were taken (Fig. 2). Sequence amplification was initially tested by comparative Southern blotting (Fig. 2A). Four individuals were analyzed: one horse (HorseS), two donkeys (DonkeyA and DonkeyB), and a mule (MuleA), offspring of DonkeyB. Signal intensity of the bands clearly indicated increased copy number of these sequences in the donkeys compared to the horse. The copy number increase is particularly marked for EAS9 and EAS18. As expected, in the mule, signal intensity was intermediate between the donkey parent and the horse sample. At the EAS19 centromeric domain, signal intensity was different in the two donkey samples, suggesting polymorphism in the population.

To quantify copy number variation, quantitative PCR (qPCR) experiments were performed, including a second horse individual (HorseT) (Fig. 2B). The results confirm sequence amplification in the two donkeys, particularly marked at the EAS9 and EAS18 centromeres (about 70- to 90-fold compared to the horses); in the mule, the copy number corresponds to about half the value of its DonkeyB father. At EAS19, the number of repeats is relatively low and differs in the two donkeys; in the mule, fold enrichment values are between those of the horses and the donkey father.

A third independent method directly compared read counts between horse and donkey input samples, aligned to the horse reference genome EquCab2.0 (Fig. 2C). The presence of peaks in the donkey centromere domains and their absence in the horse confirm that these regions are amplified in the donkey. Peak height is greater in the donkeys with respect to the mule, and the degree of amplification is lower in EAS19 compared to the other two chromosomes. Quantitative PCR experiments and input read

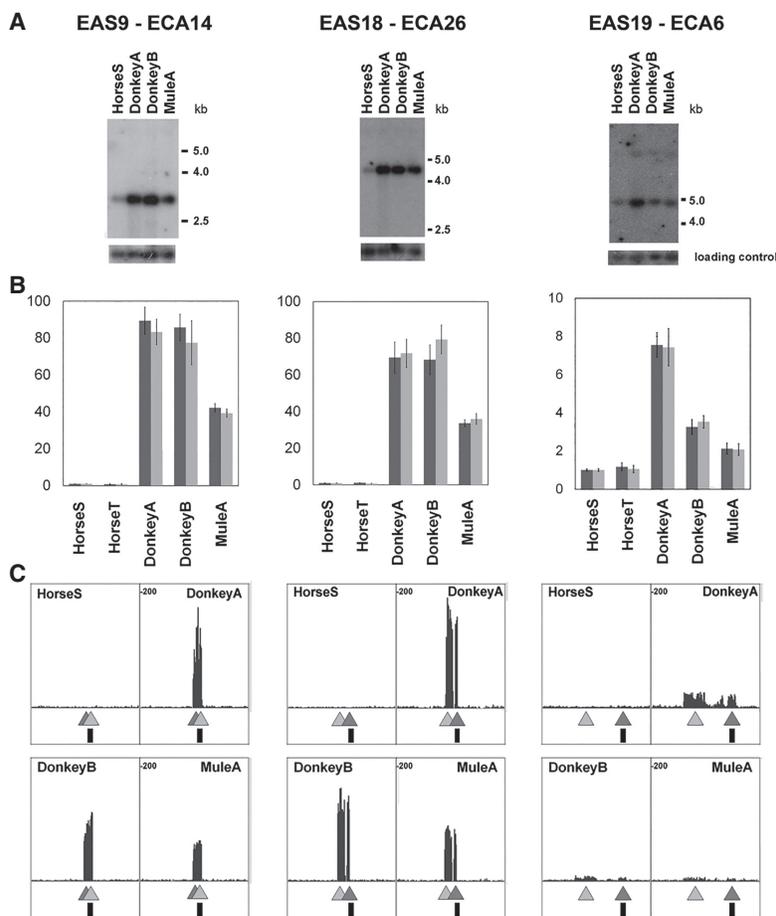


Figure 2. DNA sequence amplification at the centromeres of *E. asinus* Chromosomes 9, 18, and 19. The number of the *E. asinus* chromosome (EAS) and of its ortholog in *E. caballus* (ECA) is reported on top. (A) Southern blot analysis of genomic DNA from one horse, two donkeys, and a mule (MuleA, offspring of DonkeyB). The probes were obtained by PCR-amplification of a portion of the unit repeated in the donkey (Supplemental Table S4). Map positions of the probes are indicated as vertical black rectangles in C. (B) Quantitative PCR performed on DNA from two horses, two donkeys, and one mule. Each centromere was analyzed with two primer pairs (dark and light gray bars) (Supplemental Table S4). (C) Profile of input reads from one horse, two donkeys, and one mule aligned on the horse reference genome. The genomic regions shown are 29,593,109–29,725,206 for Chromosome 9; 22,441,448–22,572,314 for Chromosome 18; and 14,157,787–14,289,525 for Chromosome 19. Peaks represent regions amplified in the donkey genome compared to the horse genome. Light and dark gray triangles indicate the location of the fragments amplified in the quantitative PCR assay (B).

count comparisons were also carried out to analyze the variation of copy number at the centromeres of EAS16 and EAS8 (Supplemental Fig. S4), revealing sequence amplification and copy number variation.

Taken together, these results confirm the occurrence of tandem sequence amplification at a subset of centromeres in the donkey, with evidence for marked inter-individual variation in copy number at some of these loci.

DNA sequence analysis of the satellite-free centromeric domains

DNA sequence features of the satellite-free donkey centromeres were compared with the corresponding regions in the horse genome (Supplemental Fig. S5). The five centromeres containing amplifications were excluded from this analysis because we could not define their complete sequence. The percentage of SINEs, LINEs, LTR-derived sequences, and transposable DNA elements

at the donkey centromere domains did not differ from the orthologous horse sequences. The GC content at these loci was also similar in the two species. Since the horse genome sequence is not well annotated and no annotation of the donkey genome is available, we are not able to provide an accurate analysis of gene content in the satellite-free centromeric regions.

We then compared the abundance of transposable elements at the centromeric regions with the average genome-wide values obtained from a draft donkey genome (Huang et al. 2015). Donkey centromeres were significantly poor in SINES ($P < 0.00001$), whereas LINE elements were enriched ($P = 0.0057$); LTRs and DNA elements showed the same abundance in all samples. As expected, centromeric satellite sequences (Piras et al. 2010; Cerutti et al. 2016) were totally absent from the 16 centromeres examined here. Finally, donkey centromeres showed a 36.2% GC content as opposed to the genome-wide average of 41.3%, indicating that these satellite-free centromeres are AT rich.

Centromere sliding occurs in *Equus asinus*

The double peaks observed on several chromosomes (EAS5, EAS10, EAS12, EAS14, and EAS18) suggested the presence of epialleles on the homologous pairs in the donkey similarly to what we reported for horse Chromosome 11 (Purgato et al. 2015). To verify the presence of epialleles, we used a single nucleotide variant (SNV) based approach. We identified heterozygous nucleotide positions, SNVs, within each centromeric domain using a high coverage input library (Supplemental Table S1). These heterozygous positions would allow us to resolve the two homologs in the reads obtained from CENPA immunoprecipitated chromatin: If the two CENPA domains were present on both homologs, immunoprecipitated chromatin would contain similar amounts of the two SNV alleles; alternatively, if each homolog contained a single CENPA domain, only one of the two SNV alleles would be enriched in immunoprecipitated chromatin. The results of this analysis are shown in Figure 3 and Supplemental Table S5. The SNV analysis was informative for eight of the 16 centromeres (EAS4, 5, 7, 10, 12, 14, 27, and 30). The X Chromosome was excluded because this animal is a male; the five chromosomes with tandem repetitions at centromeres were excluded due to incomplete sequence definition; finally, at EAS11 and EAS13, centromeres informative SNVs were not identified. On EAS5, 10, 12, and 14 centromeres with two clearly separated peaks, a single variant was highly enriched at all positions in the immunoprecipitated DNA, demonstrating that each homolog contains a single functional domain in different positions on the two homologs (Fig. 3). On EAS4, 7, and 27, different results were obtained when SNVs at the edges or at the center of the peak

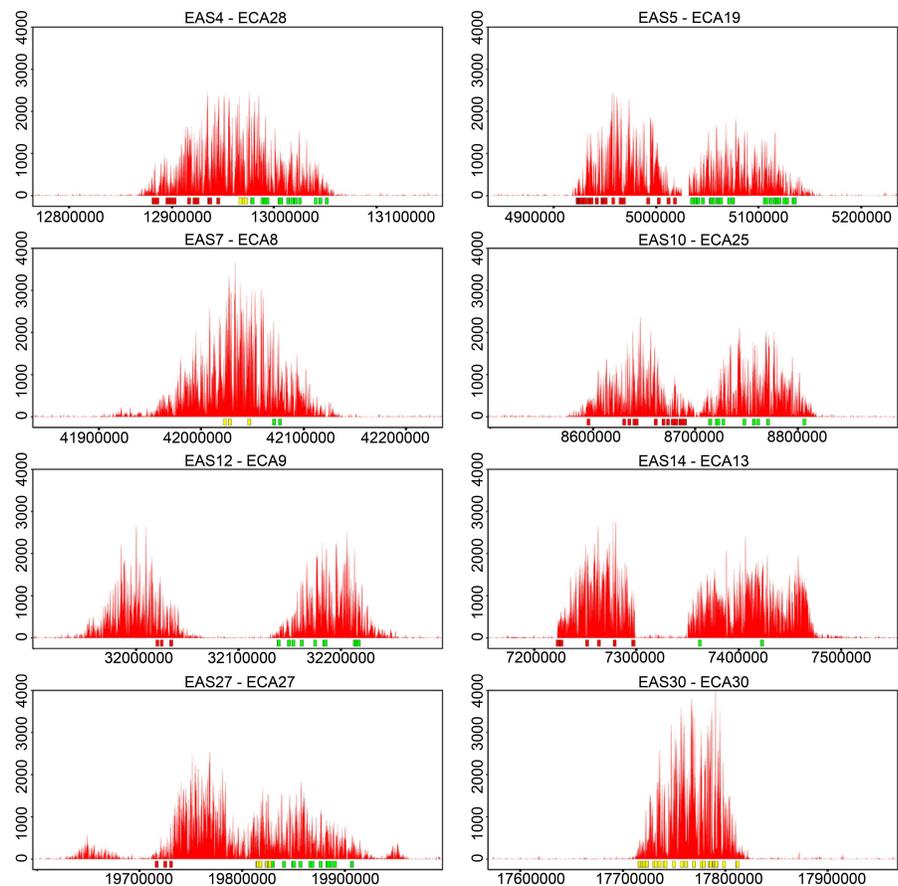


Figure 3. Identification of epialleles through SNV analysis. The positions of single nucleotide variants (SNVs), located within each centromeric domain, are represented as colored rectangles under each ChIP-seq profile. Reads were mapped on the chimeric EquCabAsiA reference genome. The y-axis reports the normalized read counts, and the x-axis reports the genomic coordinates. Red or green rectangles indicate positions where only one nucleotide variant was enriched in the immunoprecipitated reads, and yellow rectangles indicate positions where both SNVs were present.

were analyzed. At the edges, only one variant was observed; on the contrary, both nucleotides were found at the center of the peak; the interpretation of this result is that CENPA binds to slightly different but overlapping regions in the two homologs. On EAS30, at all positions both single nucleotide variants were detected, suggesting that the two homologs contain a very similar epiallele, giving rise to overlapping CENPA binding domains.

The size of individual epialleles was estimated by taking into account the borders of each peak and the distribution of SNVs (Fig. 3). This measurement is not precise, particularly when two epialleles overlap (EAS4, EAS7, and EAS27), giving rise to an approximate size of 100 kb.

To further investigate the individual variability of the donkey satellite-free centromeric domains, we analyzed an additional unrelated donkey (DonkeyB) by ChIP-seq with the same anti-CENPA antibody used for DonkeyA (Supplemental Fig. S6). To compare the two individuals, the reads of both animals were mapped on the horse reference sequence (EquCab2.0). Of the 16 satellite-free centromeres identified in DonkeyA, only 15 proved to be satellite-free in the DonkeyB: No enrichment of the ChIP-seq reads was observed on EAS8. It may be that, in DonkeyB, the centromere occurs on satellite repeats. A situation like this was recently

described in orangutan (Tolomeo et al. 2017), and we may be seeing a polymorphism in the donkey population at Chromosome 8.

A marked variability in the position of CENPA binding domains between the two individuals was observed at six chromosomes (Supplemental Figure S6), indicating that CENPA binding domains can move within regions of up to 600 kb. The remaining nine satellite-free centromeres showed little or no positional variability between these two animals.

Germline and somatic transmission of centromeric domains

The observation of positional instability of satellite-free centromeres raises the question of when such movement of the CENPA domain can occur. The stability of centromeres across generations was examined by crossing DonkeyB with three mares (HorseA, HorseB, and HorseC) by in vitro fertilization. Embryonic fibroblasts were established from the resultant mule concepti (MuleA, MuleB, and MuleC). Adult skin fibroblast cell lines were established from DonkeyB and from two of the three mares (HorseA and HorseC; cells from HorseB were not available). In addition, skin fibroblasts cell lines were obtained from a male horse (HorseD) and from the hinny derived from its cross with a female donkey (female donkey cells not available). The genetic relationships among the individuals used in this study are reported in Figure 4A. All the cell lines from the two families were subjected to ChIP-seq analysis using anti-CENPA antibody. Since the mule and hinny cells contain two haploid genomes, one from *E. caballus* and one from *E. asinus*, the transmission of individual centromere alleles could be easily followed. From the DonkeyB and the mule cell lines, three replicate ChIP-seq data sets were obtained (Methods; Supplemental Table S1).

To facilitate centromere mapping in these samples, a DonkeyB-derived chimeric genome was assembled from reads as described above for EquCabAsiA. The resulting EquCabAsiB chimeric reference sequence (Supplemental Table S3) was used to map reads deriving from DonkeyB and mule cell lines (Fig. 4B; Supplemental Fig. S7). The irregular shape of some peaks may be due to (1) inaccurate sequence assembly; (2) presence of subpopulations of cells with slightly different centromeric domains; or (3) irregular distribution of CENPA containing nucleosomes.

Figure 4B shows, as examples, the centromeric domains of Chromosomes 4 and 7 in three replicate ChIP-seq experiments carried out with the DonkeyB, MuleA, MuleB, and MuleC cell lines. The centromeres of Chromosomes 4 and 7 (Fig. 4B) showed two distinct peaks in DonkeyB, whereas each mule inherited

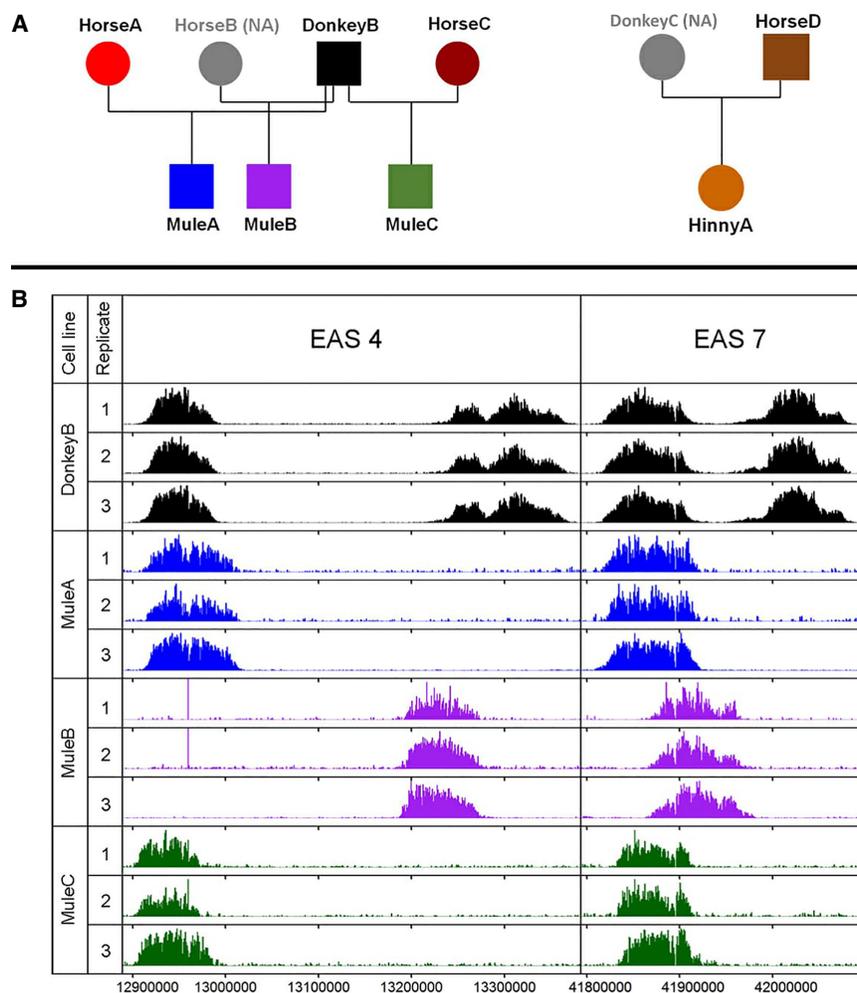


Figure 4. Transmission of satellite-free centromeric domains in hybrids. (A) Family trees reporting the genetic relationships among the individuals used in this study. Each color represents an individual, and the same color code is used in B. Cell lines from the individuals in gray were not available (NA). (B) ChIP-seq analysis performed with the anti-CENPA antibody on chromatin from the DonkeyB cell line and the cell lines from its offspring MuleA, MuleB, and MuleC. For each cell line, the results of three experiments are shown. The centromeres of donkey Chromosomes 4 (EAS4) and 7 (EAS7) are shown as examples, and the other centromeres are reported in Supplemental Figure S7. The EquCabAsiB chimeric genome was used as reference.

only one, revealing independent assortment of epialleles and normal monoallelic transmission. For Chromosome 4, the most likely interpretation is that, in MuleA, the left peak was inherited in the same position; in MuleB, the right peak was inherited but shifted by ~50 kb; and, in MuleC, the left peak was inherited with a minor, if any, movement. At Chromosome 7, the left domain seems to have been transmitted to all three mules with a relevant shift of ~50 kb in MuleB. In Supplemental Figure S7, inheritance of the other informative DonkeyB centromeric domains and of horse Chromosome 11 is shown. This analysis revealed additional examples of centromeres that exhibit a striking change in the position or structure of the epiallele in mule or hinny offspring.

In conclusion, we analyzed centromeric domain segregation of 10 donkey centromeres in three mules for a total of 30 independent events. In addition, horse Chromosome 11 centromere was analyzed in three instances. Altogether, we observed clear

positional movement in 5 of 33 transmission events. In the remaining cases, little or no movement was detected.

To test whether centromere sliding can occur during propagation in culture, we examined positional stability in six clonal cell lines isolated from TERT-TERC immortalized fibroblasts (Vidale et al. 2012) derived from MuleA. Following establishment of an immortal cell population, single cells were isolated and expanded for about 40 population doublings and subjected to CENPA ChIP-seq. As shown in Figure 5 and in Supplemental Figure S8, for 10 informative centromeres, no relevant change in peak position and shape was detected among the clones nor between the clones and the immortal parental cell line. These results suggest that the position of centromeres in the immortal cell population was homogeneous in spite of the high number of cell divisions in culture required for immortalization. In addition, during their independent growth for about 40 population doublings, centromere position remained unaltered in all the clones. In light of these observations, we can reasonably exclude in vitro cell culturing as the source of the positional instability observed in the families.

Discussion

Identification and DNA sequence composition of satellite-free centromeres

Here, we have demonstrated, at the sequence level, that an exceptionally high number of *E. asinus* centromeres are devoid of satellite DNA. If more than half of the donkey chromosomes can be stable in the species while being devoid of centromeric satellite DNA, the role of these sequences becomes even more puzzling than previously supposed (Wade et al. 2009; Fukagawa and Earnshaw 2014; Plohl et al. 2014). The 16 satellite-free donkey centromeric domains do not correspond to centromeres on the orthologous horse genomic regions; therefore, they derived from centromere repositioning events that occurred after the separation of the donkey lineage from the horse/donkey common ancestor. Thus, these centromeres are evolutionarily new (ENCs).

The large number of sequenced satellite-free centromeres allowed us to investigate the properties of “centromerizable” genomic regions in a mammal. Our analysis pointed out that satellite-free centromeres are AT and LINE rich. In addition, most satel-

lite-free centromeres contain structural rearrangements relative to *E. caballus* and, interestingly, five of 16 show sequence amplification.

Sequence analysis of the 16 satellite-free centromeric loci revealed that they are AT rich, LINE rich, and SINE poor (Supplemental Fig. S5; Huang et al. 2015). AT richness is a common feature of centromeres in a number of organisms (Clarke and Carbon 1985; Marshall et al. 2008; Chueh et al. 2009). However, it does not seem to be a necessary requirement (Melters et al. 2013), nor was it seen at the centromere of horse Chromosome 11 (Wade et al. 2009). Enrichment of LINE-1 sequences has been detected in natural human centromeres (Plohl et al. 2014) as well as in clinical neocentromeres (Chueh et al. 2005; Capozzi et al. 2008; Marshall et al. 2008). On the other hand, no association of LINES was observed in experimentally induced neocentromeres in chicken cell lines (Shang et al. 2010) or in the evolutionary neocentromere of horse Chromosome 11 (Wade et al. 2009). It is not clear whether these features contribute directly to establishment of “centromerizable” genomic domains. The observation that LINE/LTR-rich domains are clustered within the nucleus suggests that this arrangement may be related to function (van de Werken et al. 2017). In this scenario, the sequence composition of the satellite-free donkey centromeres may allow them to partition into subnuclear domains that promote the functional activation of centromeric chromatin.

Comparison between the satellite-free donkey centromeric loci and their horse noncentromeric counterparts demonstrated the presence of rearrangements in most instances (deletions, amplifications, insertions, and inversions) (Supplemental Fig. S2). Although we do not know whether these rearrangements occurred before or after centromere formation, chromosome breakage may promote CENPA binding, as suggested by the observation that CENPA can be recruited at DNA breaks (Zeitlin et al. 2009). Huang et al. (2015) used the BAC locations, mapped in our early work on centromere repositioning (Carbone et al. 2006), to identify donkey scaffolds spanning very extended regions surrounding six neocentromeres. Although they did not detect any obvious increase in chromosome rearrangements over extended (several megabases long) regions, we precisely identified sequence rearrangements contained within functional, CENPA binding, centromeric domains in this work.

Five donkey centromeres exhibit tandem repetition of sequences present in single copy in the horse genome (Fig. 2; Supplemental Figs. S2, S4). These amplified genomic sequences are unrelated to one another, with amplified units ranging in size from 5.3 (EAS16) to 138 (EAS8) kb. These repeated units are AT rich (about 65%) and SINE poor, and four of five are LINE rich. The repeat copy number was variable in the two individuals analyzed, suggesting the existence of polymorphism in the population. On the basis of our estimates, we predict that the amplified regions range in size from 100 up to 800 kb of genomic DNA. It is tempting to speculate that these amplified arrays represent an intermediate stage toward satellite DNA formation.

The presence of “ongoing” amplification at some donkey neocentromeres allows us to propose a new model (Fig. 6) for the maturation of a centromere during evolution, including different routes, some of which involve sequence amplification. According to the model, the presence of amplified sequences at a neocentromere is an indication of its more mature stage compared to nonamplified centromeres. It remains to be demonstrated whether amplification is a necessary step toward centromeric satellite DNA formation. Although the classical definition of satellite

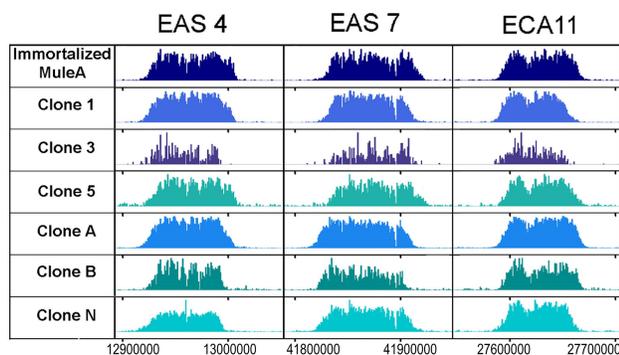


Figure 5. Transmission of satellite-free centromeric domains in clonal cell lines. ChIP-seq analysis of the immortalized cell line obtained from MuleA primary fibroblasts and six clonal derivative cell lines. Three centromeric domains taken as examples are shown (EAS4, EAS7, and ECA11). Results from the remaining centromeres are reported in Supplemental Figure S8. The EquCabAsiB chimeric genome was used as reference.

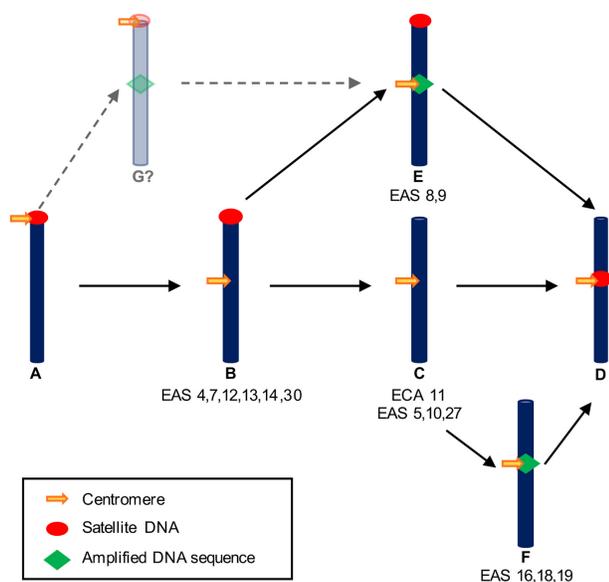


Figure 6. Model for the maturation of a centromere during evolution. Different pathways can be envisaged leading to a fully mature satellite-based repositioned centromere (D) from an ancestral centromere with satellite repeats (A) through satellite-free intermediates (B,C,E,F). The first route (A–D) follows the previously proposed model (Piras et al. 2010): a neocentromere arises in a satellite-free region; satellite repeats may then colonize this repositioned centromere at a later stage, giving rise to a “mature” centromere; meanwhile the ancestral satellite DNA is lost. Alternative routes (A, B, E, D or A, B, C, F, D) imply that, at an already functional satellite-free centromere, amplification occurs as an intermediate step toward complete maturation of the neocentromere. In this model, neocentromere maturation and loss of satellite DNA from the old centromere site are independent events that can occur at different stages during evolution. Donkey chromosomes exemplifying each step are listed, taking into account the position of satellite DNA as previously described (Piras et al. 2010). Horse Chromosome 11 is also reported since its evolutionary stage (C) was previously analyzed (Wade et al. 2009). We cannot exclude that sequence amplification may precede neocentromere formation (G?) but we have no data supporting this possibility.

DNA refers to clusters of tandem repetitions extending for several megabases, the tandem repeat expansions that we observed at these five centromeres may well be considered as an early seed of chromosome-specific centromeric satellites. In this view, these five neocentromeres cannot be considered as bona fide satellite free. To our knowledge, our results represent the first evidence supporting the hypothesis that amplification-like mechanisms can trigger the formation of tandemly repeated DNA sequences within the centromere core.

The heterogeneity of the amplified centromeric units that we observed is compatible with the molecular mechanism proposed for the multistep evolution of amplified DNA in drug-resistant mammalian cell lines (Giulotto et al. 1986). Large domains are amplified initially and, during the following steps, the copy number increases by amplification of subregions of the repeated unit, giving rise to highly condensed arrays of relatively short DNA fragments (Saito et al. 1989).

Although the systems and the time scale are extremely different, similar recombination-based mechanisms (Mondello et al. 2010) might generate novel satellite DNA families following amplification of large segments at neocentromeres. We propose that, in early stages of centromere formation, tandem duplications may arise and evolve through recombination-based meiotic or

mitotic mechanisms as demonstrated for primate alpha-satellite families (Schueler and Sullivan 2006; Cacheux et al. 2016).

In the model depicted in Figure 6, satellite DNA recruitment is a late event in centromere maturation. It has been proposed that satellite DNA increases segregation fidelity through binding with specific kinetochore proteins, such as CENPB (Fachinetti et al. 2015). The positional instability of satellite-free centromeres (discussed below) suggests that repetitive DNA arrays may contribute to centromere stability by reducing the impact of positional flexibility.

Positional variability and transmission of satellite-free centromeric domains

The position of centromeric domains can vary between individuals at satellite-free (Purgato et al. 2015) and satellite-bearing (Maloney et al. 2012) centromeres. Here, we show extensive positional allelism, verified by SNV analysis, at most donkey satellite-free centromeres (Fig. 3). Comparison of two donkey individuals (Supplemental Fig. S6) shows that centromere position can vary within genomic regions spanning several hundred kilobases, whereas independent assortment of epialleles in hybrids (Fig. 4B; Supplemental Fig. S7) provides direct proof that each chromosome carries a single centromeric domain. Despite their different positions and associated sequences, all epialleles are rather homogeneous in size, measuring ~100 kb, similar to those of horse Chromosome 11 (Purgato et al. 2015). We can reasonably propose that the sliding phenomenon is common to all satellite-free centromeres, because the analysis of only two individuals allowed us to observe evidence of more than one allele at the majority of informative centromeres (Fig. 3).

An intriguing result obtained from the analysis of the transmission of CENPA binding domains in hybrids was positional movement in five of 33 transmission events. These results demonstrate, for the first time, that centromere sliding can occur in one generation. The extent of this movement is never extreme. Indeed, the centromeric domain in the offspring is always at least partially overlapping the domain of the parent, suggesting that a fraction of CENPA nucleosomes maintains its position, and centromeres do not jump to a completely new location. We can envisage that, in the course of several generations, slight movements accumulate giving rise to nonoverlapping epialleles. In the transmission experiments reported here, we observed instances of substantial centromere movement, on the order of 50–80 kb, that occurred in a single generation. On the other hand, different epialleles at a given centromere are contained within limited regions occupying up to ~600 kb. These observations are consistent with the existence of some sort of boundaries, such as specific patterns of chromatin marks (Sullivan and Karpen 2004; Martins et al. 2016), limiting the region through which CENPA binding domains can move.

The movement of centromeric domains, observed in the family analysis, does not seem to be due to *in vitro* culturing (Fig. 5; Supplemental Fig. S8) in agreement with the behavior of centromeres in chicken DT40 cell lines (Hori et al. 2017). The stability of the centromeric domains in cultured cells is consistent with a spatially conserved transmission and replenishment mechanism for CENPA nucleosomes (McKinley and Cheeseman 2015; Ross et al. 2016) that, during the mitotic cell cycle, ensures that new CENPA nucleosomes are inserted at centromeric location with high fidelity. The sliding that we observed in the hybrids presumably took place during germline differentiation, meiotic division,

fertilization, or early developmental stages. It is possible that CENPA is mobilized during the extensive chromatin remodeling and epigenetic reprogramming characterizing these stages.

A well-described mechanism of chromatin reorganization is the replacement of histones with protamines (protamine transition) during spermatogenesis. Although CENPA is quantitatively maintained during this process (Palmer et al. 1990), it might slide into adjacent histone-depleted regions. Notably, we observed centromere sliding in both an oocyte-derived horse Chromosome 11 (Supplemental Fig. S7) as well as in several sperm-derived chromosomes in the hybrid offspring (Fig. 4B; Supplemental Fig. S7). Another process which may cause shift of centromeric domains is the meiotic division itself, during which the fidelity of CENPA deposition is poorly understood (McKinley and Cheeseman 2015). In addition, early embryonic cell cycles are highly dynamic in terms of active DNA demethylation and histone modifications and remodeling (Mayer et al. 2000; Santos et al. 2005; Probst and Almouzni 2011). We do not know at which stage centromere sliding may occur, but it is clear that the normally stringent maintenance of CENPA position can become relaxed between generations, possibly during the unique epigenetic transactions of meiosis and early embryogenesis.

Conclusions

We identified satellite-free centromeres at 16 of the 31 chromosome pairs of the donkey. Nearly one-third of the evolutionarily new centromeres of donkey exhibit tandem DNA sequence amplification. These centromeres may be in the process of selecting novel satellite DNA sequences, eventually leading to mature satellite-based centromeres (Fig. 6).

Centromeres can slide by a substantial fraction of their total size in one generation. This mobility appears to be an intrinsic property of CENPA chromatin domains in the equids. Satellite DNA may function to constrain the mobility of the centromere and enforce specific locus identity.

The presence of so many satellite-free centromeres may be due to the fact that the donkey lineage separated recently (about 3 Mya) from the common *Equus* ancestor, and there was not enough evolutionary time for satellite DNA accumulation and centromere maturation (Fig. 6). The observation of centromeres with sequence amplification intermediates supports this hypothesis. An alternative hypothesis, based on the centromere drive model (Malik and Bayes 2006; Henikoff and Furuyama 2010), can be proposed: Although large centromeres with expanded blocks of satellite DNA should be stronger than small ones (Iwata-Otsubo et al. 2017), a selective pressure against satellite DNA accumulation may operate in the donkey.

Methods

Cell lines

Primary fibroblast cell lines from HorseS and DonkeyA were established from the skin of slaughtered animals. Fibroblasts from DonkeyB, HorseA, HorseC, and Hinny were established from skin biopsies of adult animals from Cornell University. HorseD fibroblasts were obtained from testicular tissue of a freshly castrated animal from Cornell. MuleA, MuleB, and MuleC cell lines were derived from three mule conceptuses from normal pregnancies recovered on days 32–34 after ovulation via uterine lavage, as described (Adams and Antczak 2001).

Immortalization of the MuleA fibroblast cell line was carried out as described in Vidale et al. (2012) and in Supplemental Methods.

Horses, donkeys, and (horse × donkey) hybrids from the families used for the study of centromere transmission were maintained at the Baker Institute for Animal Health, College of Veterinary Medicine, Cornell University. Animal care and experiments were carried out in accordance with the guidelines set forth by the Institutional Animal Care and Use Committee of Cornell University under protocol 1986-0216, Douglas F. Antczak PI.

The DonkeyA and HorseS fibroblast cell lines were established from skin samples taken from animals not specifically sacrificed for this study; the animals were being processed as part of the normal work of the abattoirs.

Chromatin Immunoprecipitation (ChIP)

Chromatin was cross-linked with 1% formaldehyde, extracted, and sonicated to obtain DNA fragments ranging from 200 to 800 bp. Immunoprecipitation was performed as previously described (Cerutti et al. 2016) by using a polyclonal antibody against human CENPA protein (Wade et al. 2009) or a human CREST serum (Purgato et al. 2015). Sequencing was performed as described in Supplemental Methods.

Cytogenetic analysis

FISH experiments on horse and donkey metaphase spreads were carried out with a panel of BAC clones (Supplemental Table S2) from the horse library CHORI-241 as previously described (Raimondi et al. 2011; for details, see Supplemental Methods).

Assembly of centromeric regions, sequence analysis, and construction of the chimeric reference genomes

The de novo assembly of the donkey centromeric regions and the construction of the chimeric EquCabAsiA and EquCabAsiB reference genomes was performed as described in the Supplemental Methods.

Bioinformatic analysis of ChIP-seq data

Reads were aligned to the horse reference genome or to the EquCabAsiA or EquCabAsiB references with Bowtie 2.0 (Langmead and Salzberg 2012). Peak calling was performed with the software MACS 2.0.10 (Zhang et al. 2008). ChIP-seq data were normalized with the deepTools package using a subtractive method (Ramírez et al. 2014). ChIP-seq enrichment plots were obtained with the R software package Sushi (Phanstiel et al. 2014). Data sets were mapped on EquCab2.0 and plotted with Integrative Genomics Viewer (IGV) (Robinson et al. 2011). Details are reported in Supplemental Methods.

SNV analysis

To identify single nucleotide variants (SNVs) in the DonkeyA centromeric regions, we used the IGV software (Robinson et al. 2011) with the EquCabAsiA genome as reference, analyzing the BAM file resulting from read mapping (for details, see Supplemental Methods).

Southern blotting and quantitative PCR (qPCR)

Southern blotting was performed under standard conditions using probes prepared by PCR as described in Supplemental Methods.

For quantitative qPCR amplification, levels were calculated as previously described (Purgato et al. 2015). See Supplemental Methods for details.

Data access

Raw sequencing data from this study have been submitted to the NCBI BioProject database (<http://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA385275. De novo assembled centromeric regions of DonkeyA and DonkeyB from this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/genbank/>) under accession numbers MF344597–MF344627.

Acknowledgments

We thank Silvia Bione and Paolo Cremaschi (IGM-CNR, Pavia, Italy) for helpful suggestions on the initial bioinformatic analysis; Mariano Rocchi (Department of Biology, University of Bari, Italy) for providing the anti-CENPA antibody; and Claudia Alpini (Fondazione I.R.C.C.S. Policlinico San Matteo, Pavia, Italy) for the CREST serum. The E.G. laboratory was supported by grants from Consiglio Nazionale delle Ricerche (CNR-Progetto Bandiera Epigenomica, Subproject 4.9), from Ministero dell'Istruzione dell'Università e della Ricerca (MIUR): PRIN Grant No. 2015RA7XZS_002; Dipartimenti di Eccellenza Program (2018–2022) – Dept. of Biology and Biotechnology “L. Spallanzani,” University of Pavia (to S.G.N., F.M.P., M.C., E.C., E.R. and E.G.). The K.F.S. laboratory was supported by the Science Foundation Ireland under Grant No.12/A/1370. Funding bodies had no role in the design of the study and collection, analysis and interpretation of data, and in writing the manuscript.

Author contributions: E.G. conceived the study and supervised all experiments. E.G., K.F.S., and E.R. designed the research and wrote the manuscript. S.G.N., F.M.P., R.G., and M.C. carried out most molecular and cell biology experiments and bioinformatic analyses and contributed to result interpretation and figure preparation. J.G.W.McC. carried out bioinformatic analyses. Federico Cerutti, who tragically died on May 30, 2015, gave an essential contribution in the early phases of the study. E.C., F.G., R.M.H., D.F.A., D.M., M.S., and G.P. provided materials and data. D.M., R.M.H., and D.F.A. provided cells and tissues. E.G., K.F.S., E.R., S.G.N., F.M.P., R.G., M.C., F.C., J.G.W.McC., E.C., and G.P. participated in discussions and result interpretation. All authors read and approved the final manuscript.

References

Adams AP, Antczak DF. 2001. Ectopic transplantation of equine invasive trophoblast. *Biol Reprod* **64**: 753–763.

Amor DJ, Choo KH. 2002. Neocentromeres: role in human disease, evolution, and centromere study. *Am J Hum Genet* **71**: 695–714.

Cacheux L, Ponger L, Gerbault-Seureau M, Richard FA, Escudé C. 2016. Diversity and distribution of alpha satellite DNA in the genome of an Old World monkey: *Cercopithecus solatus*. *BMC Genomics* **17**: 916.

Capozzi O, Purgato S, Verdun di Cantogno L, Grosso E, Ciccone R, Zuffardi O, Della Valle G, Rocchi M. 2008. Evolutionary and clinical neocentromeres: two faces of the same coin? *Chromosoma* **117**: 339–344.

Capozzi O, Purgato S, D'Addabbo P, Archidiacono N, Battaglia P, Baroncini A, Capucci A, Stanyon R, Della Valle G, Rocchi M. 2009. Evolutionary descent of a human chromosome 6 neocentromere: a jump back to 17 million years ago. *Genome Res* **19**: 778–784.

Carbone L, Nergadze SG, Magnani E, Miscio D, Francesca Cardone M, Roberto R, Bertoni L, Attolini C, Francesca Piras M, de Jong P, et al. 2006. Evolutionary movement of centromeres in horse, donkey, and zebra. *Genomics* **87**: 777–782.

Cardone MF, Alonso A, Paziienza M, Ventura M, Montemurro G, Carbone L, de Jong PJ, Stanyon R, D'Addabbo P, Archidiacono N, et al. 2006. Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in Old World monkeys and pigs. *Genome Biol* **7**: R91.

Cerutti F, Gamba R, Mazzagatti A, Piras FM, Cappelletti E, Belloni E, Nergadze SG, Raimondi E, Giulotto E. 2016. The major horse satellite

DNA family is associated with centromere competence. *Mol Cytogenet* **9**: 35.

Chueh AC, Wong LH, Wong N, Choo KH. 2005. Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere. *Hum Mol Genet* **14**: 85–93.

Chueh AC, Northrop EL, Brettingham-Moore KH, Choo KH, Wong LH. 2009. LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. *PLoS Genet* **5**: e1000354.

Clarke L, Carbon J. 1985. The structure and function of yeast centromeres. *Annu Rev Genet* **19**: 29–55.

Cleveland DW, Mao Y, Sullivan KF. 2003. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**: 407–421.

Earnshaw WC, Migeon BR. 1985. Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma* **92**: 290–296.

Eichler EE. 1999. Repetitive conundrums of centromere structure and function. *Hum Mol Genet* **8**: 151–155.

Fachinetti D, Han JS, McMahon MA, Ly P, Abdullah A, Wong AJ, Cleveland DW. 2015. DNA sequence-specific binding of CENP-B enhances the fidelity of human centromere function. *Dev Cell* **33**: 314–327.

Ferreri GC, Liscinsky DM, Mack JA, Eldridge MD, O'Neill RJ. 2005. Retention of latent centromeres in the Mammalian genome. *J Hered* **96**: 217–224.

Fukagawa T, Earnshaw WC. 2014. The centromere: chromatin foundation for the kinetochore machinery. *Dev Cell* **30**: 496–508.

Geigl EM, Bar-David S, Beja-Pereira A, Cothran EG, Giulotto E, Hrabar H, Oyunsuren T, Pruvost M. 2016. Genetics and paleogenetics of equids. In *Wild equids* (ed. Ransom JI, Kaczensky P), pp. 87–104. Johns Hopkins University Press, Baltimore, MD.

Giulotto E, Saito I, Stark GR. 1986. Structure of DNA formed in the first step of CAD gene amplification. *EMBO J* **5**: 2115–2121.

Giulotto E, Raimondi E, Sullivan K. 2017. The unique DNA sequences underlying equine centromeres. *Prog Mol Subcell Biol* **56**: 337–354.

Gong Z, Wu Y, Kobližková A, Torres GA, Wang K, Iovene M, Neumann P, Zhang W, Novák P, Buell CR, et al. 2012. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. *Plant Cell* **24**: 3559–3574.

Han Y, Zhang Z, Liu C, Liu J, Huang S, Jiang J, Jin W. 2009. Centromere repositioning in cucurbit species: implication of the genomic impact from centromere activation and inactivation. *Proc Natl Acad Sci* **106**: 14937–14941.

Hayden KE, Strome ED, Merrett SL, Lee HR, Rudd MK, Willard HF. 2013. Sequences associated with centromere competency in the human genome. *Mol Cell Biol* **33**: 763–772.

Henikoff S, Furuyama T. 2010. Epigenetic inheritance of centromeres. *Cold Spring Harb Symp Quant Biol* **75**: 51–60.

Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**: 1098–1102.

Hori T, Kagawa N, Toyoda A, Fujiyama A, Misu S, Monma N, Makino F, Ikeo K, Fukagawa T. 2017. Constitutive centromere-associated network controls centromere drift in vertebrate cells. *J Cell Biol* **216**: 101–113.

Huang J, Zhao Y, Bai D, Shiraigol W, Li B, Yang L, Wu J, Bao W, Ren X, Jin B, et al. 2015. Donkey genome and insight into the imprinting of fast karyotype evolution. *Sci Rep* **5**: 14106.

Iwata-Otsubo A, Dawicki-McKenna JM, Akera T, Falk SJ, Chmátal L, Yang K, Sullivan BA, Schultz RM, Lampson MA, Black BE. 2017. Expanded satellite repeats amplify a discrete CENP-A nucleosome assembly site on chromosomes that drive in female meiosis. *Curr Biol* **27**: 2365–2373.e8.

Kalitsis P, Choo KA. 2012. The evolutionary life cycle of the resilient centromere. *Chromosoma* **121**: 327–340.

Karpen GH, Allshire RC. 1997. The case for epigenetic effects on centromere identity and function. *Trends Genet* **13**: 489–496.

Kobayashi T, Yamada F, Hashimoto T, Abe S, Matsuda Y, Kuroiwa A. 2008. Centromere repositioning in the X chromosome of XO/XO mammals, *Ryukyu spiny rat*. *Chromosome Res* **16**: 587–593.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Leeb T, Vogl C, Zhu B, de Jong PJ, Binns MM, Chowdhary BP, Scharfe M, Jarek M, Nordsiek G, Schrader F, et al. 2006. A human-horse comparative map based on equine BAC end sequences. *Genomics* **87**: 772–776.

Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, et al. 2011. Comparative and demographic analysis of orang-utan genomes. *Nature* **469**: 529–533.

Malik HS, Bayes JJ. 2006. Genetic conflicts during meiosis and the evolutionary origins of centromere complexity. *Biochem Soc Trans* **34**: 569–573.

Maloney KA, Sullivan LL, Matheny JE, Strome ED, Merrett SL, Ferris A, Sullivan BA. 2012. Functional epialleles at an endogenous human centromere. *Proc Natl Acad Sci* **109**: 13704–13709.

- Marshall OJ, Chueh AC, Wong LH, Choo KH. 2008. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* **82**: 261–282.
- Martins NM, Bergmann JH, Shono N, Kimura H, Larionov V, Masumoto H, Earnshaw WC. 2016. Epigenetic engineering shows that a human centromere resists silencing mediated by H3K27me3/K9me3. *Mol Biol Cell* **27**: 177–196.
- Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. 2000. Embryogenesis: demethylation of the zygotic paternal genome. *Nature* **403**: 501–502.
- McKinley KL, Cheeseman IM. 2015. The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol* **17**: 16–29.
- Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, et al. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* **14**: R10.
- Mendiburo MJ, Padeken J, Fülöp S, Schepers A, Heun P. 2011. *Drosophila* CENH3 is sufficient for centromere formation. *Science* **334**: 686–690.
- Mondello C, Smirnova A, Giulotto E. 2010. Gene amplification, radiation sensitivity and DNA double-strand breaks. *Mutat Res* **704**: 29–37.
- Montefalcone G, Tempesta S, Rocchi M, Archidiacono N. 1999. Centromere repositioning. *Genome Res* **9**: 1184–1188.
- Musilova P, Kubickova S, Vahala J, Rubes J. 2013. Subchromosomal karyotype evolution in Equidae. *Chromosome Res* **21**: 175–187.
- Nergadze SG, Belloni E, Piras FM, Khoraiuli L, Mazzagatti A, Vella F, Bensi M, Vitelli V, Giulotto E, Raimondi E. 2014. Discovery and comparative analysis of a novel satellite, EC137, in horses and other equids. *Cytogenet Genome Res* **144**: 114–123.
- Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**: 74–78.
- Palmer DK, O'Day K, Margolis RL. 1990. The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. *Chromosoma* **100**: 32–36.
- Palmer DK, O'Day K, Trong HL, Charbonneau H, Margolis RL. 1991. Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proc Natl Acad Sci* **88**: 3734–3738.
- Panchenko T, Black BE. 2009. The epigenetic basis for centromere identity. *Prog Mol Subcell Biol* **48**: 1–32.
- Phanstiel DH, Boyle AP, Araya CL, Snyder MP. 2014. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**: 2808–2810.
- Piras FM, Nergadze SG, Poletto V, Cerutti F, Ryder OA, Leeb T, Raimondi E, Giulotto E. 2009. Phylogeny of horse chromosome Sq in the genus *Equus* and centromere repositioning. *Cytogenet Genome Res* **126**: 165–172.
- Piras FM, Nergadze SG, Magnani E, Bertoni L, Attolini C, Khoraiuli L, Raimondi E, Giulotto E. 2010. Uncoupling of satellite DNA and centromeric function in the genus *Equus*. *PLoS Genet* **6**: e1000845.
- Plohl M, Meštrović N, Mravinac B. 2014. Centromere identity from the DNA point of view. *Chromosoma* **123**: 313–325.
- Probst AV, Almouzni G. 2011. Heterochromatin establishment in the context of genome-wide epigenetic reprogramming. *Trends Genet* **27**: 177–185.
- Purgato S, Belloni E, Piras FM, Zoli M, Badiale C, Cerutti F, Mazzagatti A, Perini G, Della Valle G, Nergadze SG, et al. 2015. Centromere sliding on a mammalian chromosome. *Chromosoma* **124**: 277–287.
- Raimondi E, Piras FM, Nergadze SG, Di Meo GP, Ruiz-Herrera A, Ponsà M, Ianuzzi L, Giulotto E. 2011. Polymorphic organization of constitutive heterochromatin in *Equus asinus* ($2n=62$) chromosome 1. *Hereditas* **148**: 110–113.
- Ramírez F, Düндar F, Diehl S, Grüning BA, Manke T. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187–W191.
- Raudsepp T, Lear TL, Chowdhary BP. 2002. Comparative mapping in equids: The asine X chromosome is rearranged compared to horse and Hartmann's mountain zebra. *Cytogenet Genome Res* **96**: 206–209.
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26.
- Ross JE, Woodlief KS, Sullivan BA. 2016. Inheritance of the CENP-A chromatin domain is spatially and temporally constrained at human centromeres. *Epigenetics Chromatin* **9**: 20.
- Saito I, Groves R, Giulotto E, Rolfe M, Stark GR. 1989. Evolution and stability of chromosomal DNA coamplified with the CAD gene. *Mol Cell Biol* **9**: 2445–2452.
- Santos F, Peters AH, Otte AP, Reik W, Dean W. 2005. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev Biol* **280**: 225–236.
- Schueler MG, Sullivan BA. 2006. Structural and functional dynamics of human centromeric chromatin. *Annu Rev Genomics Hum Genet* **7**: 301–313.
- Shang WH, Hori T, Toyoda A, Kato J, Pependorf K, Sakakibara Y, Fujiyama A, Fukagawa T. 2010. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res* **20**: 1219–1228.
- Steiner CC, Mittelberg A, Tursi R, Ryder OA. 2012. Molecular phylogeny of extant equids and effects of ancestral polymorphism in resolving species-level phylogenies. *Mol Phylogenet Evol* **65**: 573–581.
- Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M. 1995. A mutation in *CSE4*, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. *Genes Dev* **9**: 573–586.
- Sullivan BA, Karpen GH. 2004. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol* **11**: 1076–1083.
- Tolomeo D, Capozzi O, Stanyon R, Archidiacono N, D'Addabbo P, Catacchio CR, Purgato S, Perini G, Schempp W, Huddleston J, et al. 2017. Epigenetic origin of evolutionary novel centromeres. *Sci Rep* **7**: 41980.
- van de Werken HJ, Haan JC, Feodorova Y, Bijos D, Weuts A, Theunis K, Holwerda SJ, Meuleman W, Pagie L, Thanisch K, et al. 2017. Small chromosomal regions position themselves autonomously according to their chromatin class. *Genome Res* **27**: 922–933.
- Ventura M, Weigl S, Carbone L, Cardone MF, Misceo D, Teti M, D'Addabbo P, Wandall A, Björck E, de Jong PR, et al. 2004. Recurrent sites for new centromere seeding. *Genome Res* **14**: 1696–1703.
- Ventura M, Antonacci F, Cardone MF, Stanyon R, D'Addabbo P, Cellamare A, Sprague LJ, Eichler EE, Archidiacono N, Rocchi M. 2007. Evolutionary formation of new centromeres in macaque. *Science* **316**: 243–246.
- Vidale P, Magnani E, Nergadze SG, Santagostino M, Cristofari G, Smirnova A, Mondello C, Giulotto E. 2012. The catalytic and the RNA subunits of human telomerase are required to immortalize equid primary fibroblasts. *Chromosoma* **121**: 475–488.
- Voullaire LE, Slater HR, Petrovic V, Choo KH. 1993. A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet* **52**: 1153–1163.
- Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**: 865–867.
- Yang F, Fu B, O'Brien PC, Nie W, Ryder OA, Ferguson-Smith MA. 2004. Refined genome-wide comparative map of the domestic horse, donkey and human based on cross-species chromosome painting: insight into the occasional fertility of mules. *Chromosome Res* **12**: 65–76.
- Zeitlin SG, Baker NM, Chapados BR, Soutoglou E, Wang JY, Berns MW, Cleveland DW. 2009. Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc Natl Acad Sci* **106**: 15762–15767.
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Received October 11, 2017; accepted in revised form April 13, 2018.



Birth, evolution, and transmission of satellite-free mammalian centromeric domains

Solomon G. Nergadze, Francesca M. Piras, Riccardo Gamba, et al.

Genome Res. 2018 28: 789-799 originally published online April 30, 2018

Access the most recent version at doi:[10.1101/gr.231159.117](https://doi.org/10.1101/gr.231159.117)

Supplemental Material <http://genome.cshlp.org/content/suppl/2018/04/30/gr.231159.117.DC1>

References This article cites 77 articles, 22 of which can be accessed free at:
<http://genome.cshlp.org/content/28/6/789.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>