RESEARCH

Open Access



CENP-A/CENP-B uncoupling in the evolutionary reshuffling of centromeres in equids

Eleonora Cappelletti^{1*†}[®], Francesca M. Piras^{2*†}[®], Marialaura Biundo¹[®], Elena Raimondi¹[®], Solomon G. Nergadze¹[®] and Elena Giulotto^{1*}[®]

⁺Eleonora Cappelletti and Francesca M. Piras contributed equally to this work.

*Correspondence:

eleonora.cappelletti@unipv.it; m.piras@smatteo.pv.it; elena. giulotto@unipv.it

 Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, Pavia, Italy
Unit of Anatomic Pathology, IRCCS San Matteo Hospital Foundation, Pavia, Italy

Abstract

Background: While CENP-A is the epigenetic determinant of the centromeric function, the role of CENP-B, a centromeric protein binding a specific DNA sequence, the CENP-B-box, remains elusive. In the few mammalian species analyzed so far, the CENP-B box is contained in the major satellite repeat that is present at all centromeres, with the exception of the Y chromosome. We previously demonstrated that, in the genus *Equus*, numerous centromeres lack any satellite repeat.

Results: In four *Equus* species, CENP-B is expressed but does not bind the majority of satellite-based centromeres, or the satellite-free ones, while it is localized at several ancestral, now-inactive, centromeres. Centromeres lacking CENP-B are functional and recruit normal amounts of CENP-A and CENP-C. The absence of CENP-B is related to the lack of CENP-B boxes rather than to peculiar features of the protein itself. CENP-B boxes are present in a previously undescribed repeat which is not the major satellite bound by CENP-A. Comparative sequence analysis suggests that this satellite was centromeric in the equid ancestor, lost centromeric function during evolution, and gave rise to a shorter CENP-A bound repeat not containing the CENP-B box but enriched in dyad symmetries.

Conclusions: We propose that the uncoupling between CENP-B and CENP-A may have played a role in the extensive evolutionary reshuffling of equid centromeres. This study provides new insights into the complexity of centromere organization in a largely biodiverse world where the majority of mammalian species still have to be studied.

Keywords: CENP-B, Centromere, Satellite DNA, Karyotype evolution, Genus Equus

Background

Centromeres are essential loci required for the correct segregation of chromosomes during cell division. In higher eukaryotes, the DNA component of centromeric chromatin typically consists of tandemly repeated arrays named satellite DNA [1]. Despite the wellconserved centromeric function along the evolutionary scale, centromeric satellites are



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/. the most rapidly evolving DNA sequences in eukaryotic genomes [2–5]. According to the "library hypothesis," related species share a set of ancestral satellite families that can be differentially modified during the evolution of different lineages [6, 7]. New repeats arise and expand in the centromeric core, progressively moving the older units towards the pericentromere, forming layers of different ages [8]. During this process, pericentromeric satellites progressively become more and more degenerated and thus cannot be anymore bound by centromeric proteins, avoiding a harmful expansion of the functional centromere [9].

Although satellite DNA is usually associated to centromeres, it is neither sufficient nor necessary for specifying their function [4, 10]. Indeed, the centromeric function is not determined by the underlying DNA sequence but rather by the binding of CENP-A, a centromere-specific variant of histone H3, which is the epigenetic marker of functional centromeres [3, 10, 11].

CENP-B is highly conserved among mammals and is the sole centromeric protein so far described that exhibits unequivocal DNA binding specificity [12]. The CENP-B target site, called CENP-B box, comprises nine essential nucleotides and represents the only common motif shared by otherwise divergent centromeric satellites of different mammalian species, including several primates, rodents, marsupials and bats [13, 14]. The functional domains of CENP-B are the N-terminal DNA-binding region and the C-terminal dimerization domain which are totally conserved in primates and mouse [15, 16]. In spite of the conservation of CENP-B and its binding site, the protein is dispensable for the centromeric function. Human clinical neocentromeres and Y chromosomes from many species lack CENP-B binding sites; thus, they are not bound by CENP-B [2]. Conversely, inactive centromeres of pseudo-dicentric chromosomes can retain CENP-B, suggesting that its deposition is not sufficient for centromerization [2]. The generation of a human artificial chromosome where CENP-A chromatin was seeded on non-repetitive sequences without the requirement of CENP-B binding [17] confirmed that the absence of CENP-B is compatible with a functional centromere. Moreover, CENP-B knock-out mice are viable, mitotically and meiotically normal demonstrating that CENP-B is not essential for cell division [18-20]. These animals exhibit low body weight and uterine or testis dysfunctions suggesting a not yet known possible role of CENP-B in the physiology of the reproductive tract [18-20].

The high conservation and dispensability of CENP-B are difficult to reconcile leaving the role of this protein still controversial. It has been proposed that CENP-B might play a role in assembly, disassembly, and/or maintenance of centromere activity [21]. CENP-B may stabilize CENP-A and CENP-C maintenance at centromeres, increasing centromere strength and segregation fidelity of chromosomes [14, 22, 23]. Loss of the Y chromosome is observed in several cancer types, suggesting a high frequency of mis-segregation for this CENP-B negative chromosome [24]. It was also proposed that CENP-B participates in the formation of pericentromeric heterochromatin [25] since its depletion causes the disruption of the H3K9me3 environment around centromeres, with subsequent erosion of heterochromatin and genome instability [26, 27]. Alternatively, CENP-B conservation might be attributable to non-centromeric functions such as the silencing of transposable elements [28, 29]. Finally, it has been proposed that, in centromeric satellites harboring CENP-B boxes, CENP-B mediates the DNA bending required to adopt a non-B

conformation typically found at centromeres [30]. It has been proposed that CENP-B may collaborate with CENP-A to establish an open chromatin state by inducing nucleo-some DNA unwrapping [31].

To shed light on the role of this elusive protein, we investigated CENP-B in the genus Equus (horses, asses, and zebras) which underwent a rapid evolution after the divergence from the common ancestor, dated around 4 million years ago [32]. Asses and zebra lineages differentiated less than 1 million years ago [33-35]. The rapid evolution of these species, is marked by exceptionally frequent centromere repositioning events and chromosomal fusions that gave rise to satellite-free centromeres [33, 36-47]. In addition, blocks of satellite DNA are often present at non-centromeric chromosome ends, representing relics of ancestral inactivated centromeres or traces of satellite loci exchange [39, 46]. The chromosomal distribution of two equid satellite DNA families, 37cen and 2PI, was investigated in horse (E. caballus), donkey (E. asinus), Grevy's zebra (E. grevyi), and Burchell's zebra (*E. burchelli*) [39]. In the horse (2n = 64), all centromeres, with the exception of the one of chromosome 11 [37, 41, 44, 48], are satellite-based and the major centromeric satellite family is 37cen [39, 49]. In the donkey (2n=62), 16 satellite-free centromeres are present, while satellite DNA loci are either centromeric or non-centromeric [39, 43]. In these two species, satellite-free centromeres derive from repositioning that is the movement of the centromeric function without DNA sequence modification [50]. A high number of chromosomal fusion events led to the karyotypes of the Grevy's zebra (2n = 46) and the Burchell's zebra (2n = 44), where 13 and 15 satellite-free centromeres, respectively, were identified [46]. In the Grevy's zebra, the majority of satellite DNA loci are found at non centromeric chromosomal termini, while in the Burchell's zebra satellite DNA is mainly present at satellite-based centromeres or at fusion sites [39, 46]. Thus, the karyotypes of these species represent four different scenarios, providing the opportunity to evaluate the association between CENP-B, centromeres and satellites. Given the coexistence of satellite-free and satellite-based centromeres, the genus *Equus* is an ideal model to study the binding of CENP-B with centromeres and satellite DNA.

In this work, we analyzed the binding pattern of CENP-B in these four *Equus* species demonstrating that it is uncoupled from CENP-A binding domains. Differently from what previously observed in other systems, in our natural system, the amount of the centromeric proteins CENP-A and CENP-C is not influenced by the presence/absence of CENP-B. The CENP-B box is contained in a previously undescribed repeat that was centromeric in the equid ancestor, lost centromeric function during evolution, and gave rise to a shorter CENP-A bound repeat not containing the CENP-B box but enriched in dyad symmetries. We propose that, on an evolutionary time scale, the separation of CENP-B from CENP-A may have driven the plasticity of equid centromeres.

Results

CENP-B gene and protein conservation

The CENP-B gene sequence of horse, donkey, Grevy's zebra, and Burchell's zebra was identified in their respective genome assemblies [51, 52] (Additional file 1: Table S1) and validated using both Sanger sequencing and NGS data obtained in our laboratory (Accession Bioproject: PRJNA1054998). Comparative analysis of the DNA sequences

and of the deduced protein sequences revealed that CENP-B is highly conserved in the four species with only a few minor differences (Additional file 1: Table S1 and Fig. S1). The DNA binding and the dimerization domains are identical to the human ones (Additional file 1: Fig. S1), suggesting that the equid CENP-B is functional and able to recognize a canonical CENP-B box.

CENP-B expression was then analyzed in primary fibroblast cell lines from the four species by western blotting using an antibody against the human CENP-B protein. As shown in Fig. 1A, the protein is present in all species, although less abundant in Burchell's zebra, and, in agreement with the intracellular localization of CENP-B in human cell lines [12], resides in the nucleus.

CENP-B binding sites

We previously identified by ChIP-seq, using an anti-CENP-A antibody, one satellite-free centromere in horse [37], 16 in donkey [43], 15 in Burchell's zebra [46], and 13 in Grevy's zebra [46]. The extraordinarily high number of satellite-free centromeres in equid species raises the question whether CENP-B boxes might be present at such centromeres. We searched for CENP-B boxes (nTTCGnnnnAnnCGGGn) in the genomic sequences



Fig. 1 CENP-B protein expression and CENP-B bound satellite in horse, donkey, Grevy's and Burchell's zebra. A Left panel: western blotting on total protein extract from horse (ECA), donkey (EAS), Grevy's zebra (EGR), and Burchell's zebra (EBU) with an anti-CENP-B antibody. Protein extracts from human HeLa cells were used as control. All protein extracts were run on the same blot. Right panel: western blotting on cytoplasmic (C) and nuclear (N) protein extracts from HeLa, horse (ECA), donkey (EAS), Grevy's zebra (EGR), and Burchell's zebra (EBU) with an anti-CENP-B 07–735 antibody. Protein extracts of each species were run on different blots. B Schematic representation of the CENPB-sat satellite sequence. The CENP-B box is colored in red and the region with high identity with the 37cen satellite in yellow. C Genomic abundance of CENPB-sat in the four species. Values of genomic abundance are reported as counts per million (CPM). D In the upper row, the 9 nucleotides of the CENP-B box essential for CENP-B binding are shown. The other rows show, for each species, the consensus of the CENP-B box deduced from the Input reads

of the 45 satellite-free CENP-A binding domains [37, 41, 43, 46] of the four species and did not find any.

In all mammalian species analyzed so far, the CENP-B box is comprised within the major centromeric, CENP-A bound, satellite repeat. Surprisingly, the major horse centromeric satellite repeat that we previously identified, 37cen [49] (SAT_EC in Repbase; AY029358.1 in GenBank), does not contain any CENP-B recognition motif, and no CENP-B binding sites were detected in the 2PI satellite, the other highly represented satellite DNA family of equid species [39] (ES22 in Repbase), nor in EC137, an accessory pericentromeric satellite DNA element [53] (JX026961.1 in GenBank).

To search for CENP-B binding sites in the horse genome, we performed ChIP-seq experiments with an antibody against the human centromeric protein CENP-B on chromatin extracted from horse skin primary fibroblasts. We then aligned the reads with the horse reference genome. We did not identify any enrichment peak in the region corresponding to the satellite-free centromere of chromosome 11 (Additional file 1: Fig. S2A) while we identified several peaks in the "unplaced" genomic fraction, which includes highly repetitive DNA sequences lacking chromosomal assignment (Additional file 2: Table S2). These peaks, corresponding to CENP-B binding regions, were contained within arrays of a new satellite family, from now on termed CENPB-sat (Additional file 1: Table S3).

CENPB-sat is composed of tandem repeats of a 425 bp unit whose organization is shown in Fig. 1B. The GC content of CENPB-sat is 50.5% that is higher than the genomic average (41.0%). Each unit contains a canonical CENP-B box (5' TTTCGTCTGAGC CGGGT 3', red in the sketch of Fig. 1B) within a 201-bp fragment unrelated to any other known equine satellite (grey in Fig. 1B). The remaining 224 bps (yellow in Fig. 1B), which do not contain the CENP-B box, share 70% identity with the centromeric satellite 37cen that we previously described [49] (Additional file 1: Fig. S3A). Some CENPB-sat arrays contain degenerated CENP-B boxes or are interrupted by the 22 bp 2PI satellite (Additional file 2: Table S2).

We then carried out ChIP-seq experiments with the anti-CENP-B antibody on chromatin extracted from skin primary fibroblasts of donkey, Grevy's zebra, and Burchell's zebra. We did not identify any enrichment peak in the regions corresponding to the numerous satellite-free centromeres of these species (Additional file 1: Fig. S2B-D). We evaluated the presence and genomic abundance of CENPB-sat in the four species from the normalized number of reads in the input DNA (Fig. 1C and Additional file 1: Table S4). Grevy's zebra is the species with the highest genomic representation of CENPB-sat, followed by the horse. In donkey and Burchell's zebra, CENPB-sat is poorly represented. In Fig. 1D, the consensus of the CENP-B box in the four species is shown. These consensus sequences were deduced from the Input reads of each species aligned to the horse CENPB-sat sequence (Additional file 1: Table S3). In the horse, the CENP-B box is highly conserved. In Grevy's and Burchell's zebras, the box is well conserved and only a few mutations were observed in essential nucleotides while, in the donkey, the box is often mutated in two essential nucleotides (C4 > T and C13 > T).

We then measured the enrichment of CENPB-sat in immunoprecipitated DNA (Additional file 1: Table S4). As control, we used the ERE-1 retrotransposon, which is well conserved and interspersed throughout the equid genomes and is not expected

to be involved in the centromeric function [49, 54]. As shown in Additional file 1: Table S4, CENPB-sat is enriched in all immunoprecipitated samples, confirming that, in all species, this satellite is bound by CENP-B. The enrichment of CENPB-sat in Burchell's zebra indicates that a high fraction of the very few copies of CENPB-sat is bound by CENP-B. The low enrichment of CENPB-sat in the donkey immunoprecipitated chromatin could be due to the fact that only a fraction of the small number of its copies is bound by CENP-B, presumably because of sequence degeneration (Fig. 1D) that impairs protein recognition. To exclude that the partial identity with 37cen may bias enrichment values, we measured the enrichment of the 201-bp fragment containing the CENP-B box and not sharing any identity with 37cen (Additional file 1: Table S4). Enrichments values of the 201-bp fragment and of the entire CENPB-sat sequence were similar.

To test whether CENPB-sat is also enriched in CENP-A bound chromatin, we searched for CENPB-sat sequences in ChIP-seq reads that we previously obtained using an anti-CENP-A antibody [43, 46]. In all species, CENPB-sat is not the major centromeric satellite but only a few copies are bound by CENP-A (Additional file 1: Table S5).

A genome wide analysis of the ChIP-seq reads obtained following enrichment with anti-CENP-B antibody from the four species aligned on the horse reference genome allowed us to identify, besides CENPB-sat loci, several enrichment peaks of about 500 bp (Additional file 2: Table S6 and Table S7). These minor peaks, which did not contain any sequence matching satellite repeats, were found in all species. A subset of these peaks contained one to four CENP-B boxes or CENP-B box-like motifs (at least 7 of the 9 nucleotides essential for CENP-B binding) within single copy sequences. Several peaks mapped in the same position in different species (Additional file 2: Tables S6 and S7). Therefore, CENP-B can bind DNA sequences, not containing CENPB-sat, which are shared among different species. None of these extra-satellite peaks were located within the satellite-free centromeres that we previously described [37, 43, 46].

To test whether any CENP-B box containing satellite, other than CENPB-sat, is present in the non-caballine species, we retrieved satellite repeats using TAREAN [55] starting from our unassembled reads and did not identify any CENP-B box containing satellite other than CENPB-sat (Additional file 1: Fig. S3B). This search also revealed the presence of the previously identified satellite families (37cen, 2PI, EC137), of two novel satellite repeats (satA, satB) detected in all species, of one repeat shared by donkey and Burchell's zebra (satC) and of two repeats (satD and satE) that were detected in donkey and Burchell's zebra, respectively (Additional file 1: Table S8). The consensus sequence of the satellite families identified by TAREAN is reported in Additional file 2: Table S9. We then tested whether any of these satellite repeats were bound by CENP-A taking advantage of our previously published ChIP-seq datasets [43, 46]. Using TAREAN and ChIP-seq mapper, we confirmed that, in the horse, the most abundant satellite repeat bound by CENP-A is 37cen [49]. In the other species, where the majority of centromeres are satellite-free and satellite DNA is abundant at noncentromeric positions, the organization of the satellites bound by CENP-A is more complex (Additional file 1: Table S8). Some of the novel satellite families, such as satC in donkey and Burchell's zebra, are enriched in immunoprecipitated chromatin.

Non-canonical DNA structures

It was proposed that non-canonical DNA structures can contribute to centromere specification in the absence of CENP-B binding [30, 56]. To test whether this hypothesis may explain the peculiar relationship between CENP-B and centromeric domains in equids, we searched for dyad symmetries and other non-B forming DNA motives in the satellite-free centromeric regions of the four species using the EMBOSS palindrome and nBMST tools. For each species, we retrieved the sequence of the CENP-A binding domains [37, 41, 43, 46] and compared their content in non-B structures with those of random genomic regions with the same GC content. As shown in Additional file 1: Fig. S4, we did not detect any enrichment in these sequence features compared to random genomic regions except for A-phased repeats in Grevy's zebra. In a few cases, we detected lower levels of non-B structures in the centromeric regions. We can conclude that, in the satellite-free centromeres of these species, non-B structures are not relevant. Interestingly, when we performed the same analysis on the consensus sequences of 37cen and CENPB-sat, we observed an enrichment in the number of dyad symmetries in 37cen and in the portion of CENPB-sat sharing high sequence identity with 37cen (Additional file 1: Fig. S5).

Chromosomal localization of CENP-A, CENP-B and CENP-C proteins and of CENP-B binding satellite

Metaphase spreads from horse, donkey, Grevy's zebra and Burchell's zebra were immuno-stained with an anti-CENP-B antibody in two color immunofluorescence experiments with anti-CENP-A or anti-CENP-C antibodies. The localization of the CENP-B binding satellite (CENPB-sat) was then obtained by FISH (Fig. 2). In the four species, all primary constrictions were CENP-A and CENP-C positive, with homogeneous signal intensities, while the distribution of the CENP-B protein and of the CENPB-sat satellite was highly variable and peculiar in each species. The unexpected localization of CENP-B was confirmed using three commercial anti-CENP-B antibodies (Additional file 1: Fig. S6A) and different experimental conditions (see the "Methods" section).

In the horse, we performed immunofluorescence and FISH experiments on primary fibroblasts from two unrelated mares (Fig. 2 and Additional file 1: Fig. S6B and S7A). The same distribution of CENP-B and CENPB-sat was obtained from the two individuals. The CENP-B protein was detected at the primary constriction of nine out of the 32 chromosome pairs: three metacentric (2, 6 and 10) and six acrocentric chromosomes (17, 18, 21, 23, 24 and 29). The signal intensity of CENP-B varied greatly among different chromosomes. We could not exclude that undetectable amounts of CENP-B might be present also at some additional chromosomes. The CENPB-sat satellite could be detected at the primary constriction of five meta- or submeta-centric chromosomes (2, 3, 6, 8 and 10) and sixteen acrocentric chromosomes (14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 29, 30 and 31) (Fig. 2 and Additional file 1: Fig. S7A). All CENP-B protein signals colocalized with CENPB-sat signals while, on 12 centromeres, we could detect CENPB-sat signals only (3, 8, 14, 15, 16, 19, 20, 22, 25, 27, 30 and 31) (Fig. 2). The lack of detectable CENP-B protein signals at a subset of CENPB-sat positive loci was confirmed by immuno-FISH experiments (Fig. 3A). It



Fig. 2 Localization of CENP-B and CENPB-sat in the four species. First column: double immunofluorescence with an anti-CENP-B antibody (red) and an anti-CENP-A serum (green) on DAPI-stained metaphase chromosomes (blue). Second column: double immunofluorescence with an anti-CENP-B antibody (red) and an anti-CENP-C serum (green) on DAPI-stained metaphase chromosomes. Third column: FISH localization of CENPB-sat (red) on DAPI-stained metaphase chromosomes. Fourth column: schematic representation of CENP-B and CENPB-sat signals on metaphase chromosomes. Loci hybridizing with the CENPB-sat probe only are labeled in red. Loci hybridizing with the CENPB-sat probe and positive to CENP-B immunofluorescence are labeled in yellow

is likely that sequence degeneration of the CENP-B box, not detectable by FISH, may prevent binding of the protein at these loci. We cannot exclude that small amounts of CENP-B protein were present at these loci but were undetectable due to the low resolution of the technique.

To confirm the distribution of the CENP-B protein on horse chromosomes, an alternative cellular system, not based on the use of antibodies, was set up. A horse fibroblast cell line, previously immortalized in our laboratory by human telomerase overexpression [57], was transfected with a construct containing the horse *CENP-B* gene tagged with eGFP. As shown in Fig. 3B, only a subset of chromosomes were eGFP positive while most chromosomes lacked detectable eGFP signals. We counted the number of eGFP signals in 10 metaphase spreads. Between fourteen and eighteen signals per metaphase were counted; therefore, the distribution of the chimeric protein in transfected cells confirmed the results obtained with the antibodies.

In the donkey, the analysis of two unrelated individuals showed that, although all centromeres were labeled by CENP-A and CENP-C, no CENP-B signal could be detected (Fig. 2, Additional file 1: Fig. S6B). This observation indicates that not only the 16 satellite free centromeres but also the 16 satellite-based centromeres are not bound by detectable levels of CENP-B. FISH experiments with the CENPB-sat probe showed hybridization signals on the primary constriction of chromosome 3



Fig. 3 CENP-B binding in horse, Grevy's zebra, and mule. A Localization of CENP-B protein and CENPB-sat on horse and Grevy's zebra metaphase chromosomes by immuno-FISH. Left: CENP-B signals (green) on DAPI-stained chromosomes (blue). Middle: FISH CENPB-sat signals (red) on the same metaphase spreads. Immunofluorescence and FISH signals were acquired separately. White arrows point to examples of chromosomes with CENPB-sat but without CENP-B signals. Right: DAPI staining of the same chromosomes. B A schematic representation of the CENPB-eGFP construct used in transfection is shown on top. Detection of eGFP tagged CENP-B (green) in three horse metaphase spreads. A CENP-B positive and two CENP-B negative chromosomes are boxed and zoomed in the top panel. C Localization of CENP-A and CENP-B in mule primary fibroblasts. Immunofluorescence with an anti-CENP-A serum (green) and an anti-CENP-B antibody (red) on DAPI-stained metaphase chromosomes

only (Fig. 2 Additional file 1: Fig. S7B). However, as in several horse centromeres, the CENP-B protein was not detected on this CENPB-sat positive chromosome.

To confirm the lack of CENP-B signal on donkey chromosomes, we examined CENP-B localization in a fibroblast cell line from a mule that is a hybrid between a horse mare and a donkey jack. Following a double immunofluorescence experiment on metaphase spreads with anti-CENP-B and anti-CENP-A antibodies, specific CENP-B signals were detected only on the primary constriction of the nine chromosomes previously identified in the horse while all other chromosomes, including the complete donkey set, were not labeled (Fig. 3C). Thus, we confirmed that, also in a cell line from a mule, no CENP-B protein signal could be detected on the donkey chromosomes. These results indicate that the absence of CENP-B protein binding at donkey centromeres is related to the lack of CENP-B boxes rather than to peculiar features of the protein itself.

In a fibroblast cell line from a Grevy's zebra female, CENP-B protein signals were detected at fifteen loci. Surprisingly, only two signals were localized at primary constrictions (chromosomes 7 and 12) while the remaining thirteen were at non-centromeric termini (Fig. 2). In particular, CENP-B localized at a non-centromeric terminus of ten meta- or sub-metacentric (1p, 2p, 5p, 7p, 10p, 12p, 13p, 14p, 15p and 16p) and three acrocentric chromosomes (20g, 21g and 22g). The CENPB-sat and the CENP-B protein colocalized at all sites with the exception of the termini of chromosomes 6, 8, and 19 where only the satellite signal was detected (Fig. 2). The results of immuno-FISH experiments confirm this observation (Fig. 3A). On chromosomes 2, 13, 16, and 19 the terminal non centromeric signals of CENPB-sat and CENP-B showed different intensities on the two homologs suggesting that polymorphism in the copy number of the CENPB-sat repeats may be present in the population (Additional file 1: Fig. S8). The localization of CENPB-sat was then analyzed in a fibroblast cell line from a male Grevy's zebra unrelated to the female of Fig. 2. As shown in Additional file 1: Fig. S7C, in the second individual, the distribution of CENPB-sat is similar to that of the first individual but signal heterogeneity in additional homologous chromosomes was observed (Additional file 1: Fig. S7C), confirming the variability of extra-centromeric CENPB-sat loci in the population.

In Burchell's zebra, as in the donkey, no CENP-B signals could be detected, whereas CENP-A and CENP-C signals were homogeneous on all primary constrictions (Fig. 2). Accordingly, no CENPB-sat hybridization signals were detected in this species, confirming the results of genome sequence analysis (Fig. 1C).

We then performed 3D-immunofluorescence experiments using anti-CENP-B and anti-tubulin antibodies (Fig. 4, Additional file 3–10: movies S1-S8). With this methodology cell morphology is preserved as opposed to the method used to prepare metaphase spreads (Fig. 2). Since it is well known that CENP-B localizes at all human centromeres, HeLa cells were used as control. As expected, in HeLa cells, CENP-B fluorescence was present in the nucleus, with discrete foci corresponding to centromeres (Fig. 4A, Additional file 3: movie S1). In horse and Grevy's zebra, the situation was similar, with discrete nuclear CENP-B foci (Fig. 4A, Additional files 4–6: movies S2-S4). On the contrary, donkey and Burchell's zebra lacked CENP-B nuclear foci and only a diffuse fluorescence was observed (Fig. 4A, Additional files 7–10:



Fig. 4 Localization of CENP-B in interphase and metaphase cells by 3D immunofluorescence. **A** Optical sections from 3D-immunofluorescence with anti-CENP-B (red) and an anti-tubulin (green) antibodies on whole cells in HeLa, horse, donkey, Grevy's zebra, and Burchell's zebra. **B** Optical sections from 3D-immunofluorescence with an anti-CENP-B antibody (red) on metaphase cells in HeLa, horse, donkey, Grevy's zebra, and Burchell's zebra. **Nuclei** were counterstained with DAPI (blue). Bars = 10 µm

movies S5-S8). This result is consistent with the absence of chromosomal CENP-B loci detectable by immunofluorescence (Fig. 2). The presence of discrete chromosomal CENP-B loci only in horse and Grevy's zebra was also observed in metaphase cells (Fig. 4B, Additional files 5–6: movie S3 and S4) confirming the results obtained with metaphase spreads (Fig. 2).

CENP-B positive and negative chromosomes: CENP-A and CENP-C quantification and segregation fidelity

It has been proposed that the amount of CENP-B protein at centromeres directly correlates with the amount of CENP-C, resulting in different degrees of centromere stability [14].

Taking advantage of the presence of CENP-B positive and CENP-B negative centromeres in the horse, we tested the possible correlation among the levels of the three centromeric proteins by immunofluorescence. As shown in Fig. 5A and B, the centromeric CENP-B signals did not show the typical speckled pattern of CENP-A and CENP-C but were broad extending over the pericentromeric area, confirming that CENP-B is localized outside the centromeric core. This figure clearly shows that the intensity of CENP-A and CENP-C signals is homogeneous regardless the presence or absence of CENP-B signals. We then measured fluorescence intensities of CENP-A, CENP-B, and CENP-C signals using the program Fiji. As shown in Fig. 5C and D (left panels), CENP-A and CENP-C signals did not differ in CENP-B positive and negative centromeres according to Student's *t* test (*t*-value = 1.479 and *p*-value = 0.14 for CENP-A; *t*-value = 0.898 and *p*-value = 0.37 for CENP-C) (Additional file 1: Table S10). The fluorescence intensity of CENP-B signals was then plotted against the intensity of CENP-A or CENP-C signals (Fig. 5C and D, right panels). According to Spearman's correlation test, fluorescence intensity of CENP-B signals was not correlated with signal intensity of CENP-A

(See figure on next page.)

Fig. 5 CENP-A and CENP-C localization and segregation fidelity on CENP-B positive and negative chromosomes. A Examples of chromosomes immuno-stained with anti-CENP-B antibody (red) and anti-CENP-A serum (green). CENP-A signals are homogeneous both in CENP-B positive and negative centromeres. B Examples of chromosomes immuno-stained with anti-CENP-B antibody (red) and anti-CENP-C serum (green). CENP-C signals are homogeneous both in CENP-B positive and negative centromeres. C CENP-A and CENP-B fluorescence intensity. Left: CENP-A fluorescence intensity in CENP-B positive (red), CENP-B negative (yellow) and all (orange) centromeres. ns: not significant. Right: absence of correlation between CENP-A and CENP-B fluorescence intensity. Each dot corresponds to a centromere and each color corresponds to a metaphase spread. Statistics are reported in Table S10. D CENP-C and CENP-B fluorescence intensity. This panel is structured as panel C. E Mitotic stability of ECA9 (CENP-B negative) and ECA10 (CENP-B positive) chromosomes by interphase aneuploidy analysis. Chromosome-specific BAC probes were used in FISH experiments, and the number of signals per nucleus was counted in two independent experiments. Nuclei with one or three signals were considered aneuploid. In the second experiment, the number of aneuploid nuclei was counted both in normal conditions and following mitotic stress induced by a 48-h treatment with 200 nM nocodazole. The numbers of counted nuclei are reported in Table S11. F Fraction of CENP-B positive chromosomes in horse immortalized fibroblasts. Left: fraction of CENP-B positive chromosomes per metaphase spread across three different passages in culture. ns: not significant. Middle: fraction of CENP-B positive chromosomes against chromosome number. Right: number of CENP-B negative chromosomes against chromosome number. Statistics is reported in Table S12. Colors correspond to passage number. G Fraction of CENP-B positive chromosomes in mule immortalized fibroblasts at different passages in culture. This panel is structured as panel F



(rho = 0.07 and *p*-value = 0.216) and CENP-C (rho = 0.03 and *p*-value = 0.616 for CENP-C) (Additional file 1: Table S10). Thus, differently from previous results in human and mouse [14], the levels of CENP-A and CENP-C were rather homogeneous and independent from the presence and amount of CENP-B.

We then compared the mitotic stability of a CENP-B positive (ECA10) and a CENP-B negative (ECA9) chromosome by interphase aneuploidy analysis (Fig. 5E). Chromosome-specific BAC probes were used in FISH experiments, and the number of signals per nucleus was counted in two independent experiments. The numbers of counted nuclei are reported in Additional file 1: Table S11. Nuclei with one or three signals were considered aneuploid. In the second experiment, the number of aneuploid nuclei was counted both in normal conditions and following mitotic stress induced by the spindle inhibitor nocodazole. The results showed that segregation fidelity was not influenced by CENP-B.

Since equid primary fibroblasts have a lifespan of about 20 passages before senescence, we utilized horse and mule fibroblast cell lines immortalized with telomerase [43, 58] to better determine the long-term segregation dynamics of chromosomes relative to the amount of centromeric CENP-B. As previously shown in human fibroblasts immortalized by telomerase [59], the chromosome number of these cell lines increased at late passages (Fig. 5F and G). In the horse cell line, the fraction of CENP-B positive chromosomes was maintained during long-term culturing (Fig. 5F). In the mule cell line, the fraction of CENP-B positive chromosomes decreased during long term culturing (Fig. 5G). These results suggest that, in these experimental conditions, the presence of CENP-B does not offer any selective advantage in chromosome segregation during long-term culture.

CENP-B binding satellite and karyotype evolution

We carried out a comparative analysis of the position of CENPB-sat loci between horse and Grevy's zebra (Fig. 6). To construct this figure, we performed a whole-genome alignment of the horse and the Grevy's zebra genome assemblies (Additional file 1: Fig. S9). A detailed description of the comparative analysis is reported in the Supplementary text. Briefly, we observed four different situations: (1) maintenance of the localization of CENPB-sat at horse centromeres and at orthologous centromeric (EGR7cen/ECA15cen and EGR12cen/ECA20cen) or terminal (EGR7pter/ECA2cen, EGR8pter/ECA31cen, EGR12pter/ECA8cen and EGR16pter/ECA24cen) positions in the zebra. The terminal zebra positions can be interpreted as remnants of ancient centromeres that were inactivated in the zebra and conserved in the horse; (2) loss of CENPB-sat in the zebra compared to orthologous centromeric horse positions (EGR1/ECA25cen-ECA16cen, EGR3/ECA2cen-ECA3cen, EGR5/ECA14cen, EGR6/ECA17cen, EGR9/ECA22cen-ECA18cen, EGR11/ECA21cen-ECA19cen, EGR14/ECA6cen, EGR17/ECA3cen-ECA10cen, and EGR18/ECA8cen) following Robertsonian fusion or other rearrangements; (3) presence of CENPB-sat on a non-centromeric terminus of zebra chromosomes and absence on the horse orthologous chromosome (EGR2pter/ECA1pter, EGR5pter/ECA13qter, EGR13pter/ECA5pter, EGR14pter/ECA12qter, EGR15pter/ECA9pter, and EGR20qter/ ECA26qter). Terminal CENPB-sat zebra positions may correspond to ancestral centromeres that were inactivated in the horse; (4) presence of CENPB-sat at terminal zebra positions and at the opposite centromeric end in the horse orthologous chromosome (EGR1pter/ECA6cen, EGR6pter/ECA23cen, EGR10pter/ECA10cen, EGR19qter/ECA-27cen, EGR21qter/ECA29cen, and EGR22qter/ECA30cen). This peculiar comparative localization is likely a consequence of satellite DNA exchange between opposite chromosomal termini [60, 61].



Fig. 6 Comparison between Grevy's zebra and orthologous horse chromosomes. Colors refer to orthologous sequences. Inverted segments are indicated with crossed lines. The position of centromeres (white ovals), CENPB-sat (red lozenges) and other satellite families (yellow lozenges) are indicated. A detailed description of this figure is reported in Supplementary text

Discussion

In previous work, we discovered that the great karyotype heterogeneity of the otherwise closely related *Equus* species is mainly due to centromere movements that occurred during evolution through either centromere repositioning or chromosome fusion. This extensive reshuffling generated numerous satellite-free centromeres. The main question arising from these previous observations was: what makes *Equus* centromeres so plastic compared to those of the other mammalian species studied so far? In the present work, we report on a peculiarity of CENP-B binding pattern and on its dissociation from CENP-A.

The first indication that CENP-B is not associated to CENP-A came from the discovery that the 45 satellite-free centromeres that we identified in four *Equus* species do not contain any CENP-B binding motif and are not bound by CENP-B. The CENP-B box is also missing in the main CENP-A bound satellite of these species but is contained in a novel satellite, CENPB-sat, which is mainly pericentromeric or located at ancestral inactivated centromeres. CENPB-sat is composed of tandemly repeated 425 bp monomers, arranged in a head-to-tail fashion. While the majority of centromeric satellites are AT rich [62], CENPB-sat is GC rich. Interestingly, we previously showed that also the major horse CENP-A binding satellite, 37cen, is GC rich [49]. A 224 bp fragment of CENPBsat, which does not contain the CENP-B box, shares 70% identity with 37cen suggesting a common evolutionary origin for CENPB-sat and 37cen.

In human and mouse, the CENP-B protein is localized at all centromeres due to the presence of CENP-B boxes within the centromeric satellites of these species. It has been stunning to observe, in the equids, a completely different binding pattern of CENP-B, which is often uncoupled from primary constrictions. In the horse, only 9 primary constrictions are bound by CENP-B while, at the unique satellite-free centromere of chromosome 11 and at 22 of the 31 satellite-based centromeres, no CENP-B binding was detected. On the other hand, the CENPB-sat satellite was detected cytogenetically at 21 primary constrictions, suggesting that, at several loci, this sequence underwent degeneration losing the ability to be recognized by the protein.

In donkey and Burchell's zebra, CENP-B was not detected at any chromosome and the genomic amounts of CENPB-sat were extremely low. In the donkey, the consensus sequence of the CENP-B box obtained from the very few copies of CENPB-sat revealed frequent mutation of two essential nucleotides. The low enrichment of CENPB-sat in donkey CENP-B bound chromatin is another evidence of CENPB-sat degeneration. Therefore, both reduction and degeneration of binding sites are responsible for the absence of detectable levels of CENP-B. In Burchell' zebra, the very few copies of CENPB-sat contain a canonical CENP-B box, therefore the lack of detectable CENP-B protein binding is due to the extreme paucity of binding sites.

In Grevy's zebra the CENP-B protein was detected at two primary constrictions only and at one non-centromeric end of 13 out of the 23 chromosomes. To our knowledge, this is the first report of such extreme uncoupling between CENP-B and centromeric function. The great abundance of CENPB-sat in this species is mainly due to its localization within satellite arrays at chromosomal termini.

A situation in which only a subset of centromeres shows levels of CENP-B binding detectable by immunofluorescence was previously observed in some New World monkeys [63, 64]. In other monkeys, such as the African Green Monkey, the amount of CENP-B bound to centromeres was lower compared to other primate species, including humans, due to low abundance of CENP-B boxes [65, 66]. However, in these monkeys, CENP-B boxes are still contained in the centromeric alpha satellite family and the absence or reduction of CENP-B binding was thought to reflect the presence of monomers defective in CENP-B binding sites. Differently, in equids, CENP-B boxes are confined to a satellite family which is not enriched in the centromeric core.

An intriguing finding was the presence of CENP-B enrichment peaks, identified by ChIP-seq, at intrachromosomal non-satellite positions. Several sites are shared among different species and only a subset of them contains CENP-B box-like motifs. These results suggest that CENP-B can bind DNA sequences other than the CENP-B box possibly exerting additional functions unrelated to centromeres.

In human and mouse experimental systems, where CENP-A was perturbed, the amounts of CENP-B and CENP-C were correlated and reduced levels of CENP-B seemed to be associated to an increased frequency of mis-segregation [14, 22, 23]. At a human neocentromere and at the centromere of human chromosome Y, which are devoid of CENP-B, similar results were obtained under normal conditions [22]. These experiments showed that, at these CENP-B negative centromeres, the amount of CENP-C was about half of that of the other chromosomes resulting in an increased mis-segregation frequency. On the contrary, in our natural system, the recruitment of CENP-A and CENP-C was not related to the amount of CENP-B. Indeed, while the amount of CENP-B was highly heterogeneous among different chromosomes and no detectable CENP-B was observed at several centromeres, the levels of CENP-A and CENP-C were homogeneous. These findings underline the difference between the human and the equid system and suggest that the interaction among CENP-B, CENP-A, and CENP-C might be more complex than previously proposed.

In previous work, we compared the mitotic stability of horse chromosome 11, whose centromere is satellite-free, with horse chromosome 13, whose centromere is satellite based but, as we know now, not bound by CENP-B. We demonstrated that segregation fidelity was not influenced by the presence of satellite DNA at the centromere [67]. In the present work, we compared the frequency of nuclei aneuploid for the CENP-B positive chromosome 10 and the CENP-B negative chromosome 9 demonstrating that mitotic segregation fidelity was not affected by the absence of CENP-B. The analysis of long-term segregation dynamics of CENP-B positive and negative chromosomes in immortalized horse and mule cell lines revealed that, during long-term culture leading to hyperdiploid karyotypes, CENP-B positive chromosomes do not have a selective advantage over CENP-B negative chromosomes. The reduced fraction of CENP-B positive chromosomes in mule cells at late passages may be due to a selective advantage, in this culture conditions, of some CENP-B negative chromosomes and/or to random mis-segregation. These results are not surprising considering that CENP-B negative centromeres are fixed in the *Equus* populations and that these populations are composed by healthy, normally developing fertile individuals.

On the basis of the cytogenetic and ChIP-seq data presented in our previous [39, 49] and present work, we propose the model depicted in Fig. 7 to interpret the evolution of CENPB-sat in the *Equus* species. According to this model, centromeres with CENPB-sat



Fig. 7 Model for CENPB-sat and 37cen evolution. The different organizations of satellites are sketched over a line representing the genomic position and CENP-A binding (yellow bars). Chromosome numbers displaying each configuration are listed under each sketch. In the equid ancestor, the centromeric CENP-A binding domains were constituted by arrays of the CENPB-sat satellite which contained a functional CENP-B box and was bound by CENP-B (red circles). Subsequently, 37cen arrays were generated by the expansion of the portion of CENPB-sat not containing the CENP-B box pushing entire CENPB-sat units outwards and colonizing the CENP-A binding domain. During the evolution of the horse lineage, at some centromeres, CENPB-sat repeats lost the ability to bind CENP-B due to mutations in the CENP-B box while at other centromeres only 37cen arrays were maintained and bound by CENP-A. In this lineage, the 37cen satellite became the major CENP-A binding centromeric satellite. In the donkey lineage, where most centromeres are satellite-free, degenerated CENPB-sat arrays were maintained at chromosome 3 only. 37cen arrays were mainly kept at non-centromeric chromosome ends (blue bars) corresponding to ancestral inactivated centromeres. In the Grevy's zebra, the two ancestral centromere configurations can be still observed together with conserved or degenerated CENPB-sat arrays at non-centromeric chromosome ends, corresponding to inactivated centromeres. In the Burchell's zebra, both CENPB-sat and 37cen repeats are nearly absent

repeats and CENP-B binding correspond to the ancestral configuration of Equus centromeres which is maintained at horse chromosome 2 and Grevy's zebra chromosome 12. In a common ancestor of all *Equus* species, the expansion of the portion of CENPBsat lacking the CENP-B box gave rise to arrays of 37cen where the functional CENP-A binding centromere was seeded, whereas the CENP-B binding repeats were pushed towards the pericentromeric regions. Indeed, new satellite sequences are known to arise and expand in the centromeric core, progressively moving the older units towards the pericentromere, forming layers of different ages [8, 68, 69]. It was proposed that pericentromeric satellites progressively become more and more degenerated and thus cannot be bound anymore by centromeric proteins, avoiding a harmful expansion of the functional centromere [9]. In agreement with this view, most horse CENPB-sat loci identified by FISH are presumably degenerated and therefore no more able to bind CENP-B (Fig. 2). The presence of CENPB-sat at most horse acrocentric chromosomes (Fig. 2) further supports this hypothesis since, as mentioned above, these chromosomes correspond to ancestral ones [34, 40]. Another evidence that CENPB-sat is an ancestral satellite and 37cen emerged in relatively recent evolutionary times is given by the fact that those horse metacentric chromosomes which are evolutionarily recent and derive from centromere repositioning or inversion (ECA1, ECA4, ECA7, ECA9, ECA11, ECA12, ECA13) [33] lack CENPB-sat and contain 37cen arrays. A possible explanation of this observation is that these centromeres were born satellite-free and, with the exception of the ECA11 centromere, progressively accumulated 37cen repeats during their maturation [37, 39, 43, 49]. A similar situation was described in primates where centromeres deriving from centromere repositioning have acquired satellite repeats during their evolutionary maturation [50, 61].

In the donkey, no binding of CENP-B was detectable and most of the few copies of the CENP-B box are degenerated. A faint CENPB-sat FISH signal at the primary constriction of chromosome 3 suggests the presence of degenerated repeats at this locus. Arrays of the 37cen satellite were observed at two donkey centromeres and at several non-centromeric chromosome ends corresponding to ancestral inactivated centromeres (Fig. 7) [39, 43].

In Grevy's zebra, contrary to horse and donkey, the majority of CENPB-sat loci are present at terminal non-centromeric positions as relics of ancestral inactivated centromeres. The copy number of some non-centromeric repeats varied in the two individuals and between homologous chromosomes of each individual suggesting that these loci are polymorphic in the population. This variability would imply progressive loss of CENP-B binding sites due to the uncoupling between CENP-B and the centromere. The 37cen sequence was detected at one primary constriction only (EGR7) [39], which contains also arrays of CENPB-sat and is bound by CENP-B. The extended and conserved CENPB-sat arrays of the Grevy's zebra at non centromeric termini could represent relics of ancestral inactivated centromeres suggesting that this species might be closer to the common ancestor than asses and other zebras. Indeed, *Equus grevyi* is the only extant member of the subgenus *Dolichohippus* and, according to paleontological, ecological, and morphological evidence, is considered closer to the Eurasian ancestor than the other zebras, which are grouped in the subgenus *Hippotigris* [70, 71].

In Burchell's zebra, we could not identify any CENP-B or CENPB-sat positive chromosome by immunofluorescence and FISH, in agreement with the extreme paucity of CENP-B binding sites revealed by sequencing. No 37cen loci were detected as well [39]. In this species, the organization of satellite DNA is relatively dynamic due to the presence of novel repeats that are probably evolutionarily recent.

In all the four species, the 2PI satellite, which is not enriched in CENP-A chromatin, is one of the most abundant satellite families. This satellite is found at most horse primary constrictions, at numerous donkey and Grevy's zebra non centromeric termini and non-centromeric interstitial or terminal positions of Burchell's zebra [39, 46]. This distribution suggests that the 2PI observed now may be the relics of the oldest equid centromeric satellite that was progressively dissociated from the centromeric function following the expansion of CENPB-sat and 37cen satellites in a lineage-specific manner. This hypothesis is supported by the variability of 2PI units (Additional file 1: Table S8 and Additional file 2: Table S9).

According to the model shown in Fig. 7, in the common ancestor of equids, centromeric DNA was composed by arrays of CENPB-sat containing functional CENP-B boxes and thus binding CENP-B. A key question then arises: why, during the evolution of equids, did the centromere function escape from CENP-B binding satellite repeats landing either into arrays not containing any CENP-B box or into satellite free regions? We, and other authors [2, 36, 42, 43, 45, 46, 50, 72], hypothesized that the loss or rearrangement of centromeric satellite repeats may have triggered the movement of the centromeric function to a new "centromerizable" position that was favored by the epigenetic context. In this scenario, the absence of satellite DNA may be the result of selective pressure favoring epigenetic factors rather than DNA sequences. A factor possibly contributing to centromere reshuffling is DNA methylation. It has been shown that both CENP-A and CENP-B preferentially bind regions of reduced CpG methylation [73, 74]. Since the methylation of the CpG dinucleotides in the CENP-B box motif is known to prevent CENP-B binding [73], both sequence degeneration of the box and/or changes in its methylation status might have contributed to the loss of CENP-A and CENP-B binding at numerous CENPB-sat loci. Another explanation of the absence of CENPB boxes in the centromeric satellite repeats of equids is that a 200-bp deletion in the ancestral CENP-B box containing repeat may have generated a variant that maintained a favorable secondary structure for CENP-A binding while pushing away the original repeat that became pericentromeric. It has been proposed that noncanonical DNA structures may contribute to centromere specification. These peculiar configurations may arise in the presence of sequence features, such as dyad symmetries or non-B DNA forming motifs, or thanks to the bending activity of sequence-specific DNA-binding proteins such as CENP-B [30]. The 37cen horse sequence, as already mentioned by Kasinathan and Henikoff [30], is enriched in dyad symmetries which facilitate the adoption of stable secondary structures. In the present work, we found that, in the CENPB-sat sequence, dyad symmetries are restricted to the portion sharing identity with 37cen, suggesting that the expansion of 37cen satellite in the centromeric core could be favored, replacing the arrays of the entire CENPB-sat sequence in the centromeric cores. Another factor promoting the expansion of 37cen may be its length that is about half that of CENPB-sat and possibly easier to be phased with nucleosome wrapping [75-77]. However, we did not find any non-B motif enrichment in the satellite-free centromeres suggesting that epigenetic factors such as heterochromatic histone marks or alterations in DNA methylation patterns may contribute to centromere specification.

In conclusion, CENP-B has been proposed to be involved in centromere strength and stability [22, 23, 78] and maintenance of pericentric heterochromatin, acting as a barrier against genome instability [26, 27]. Interestingly, in donkey, Grevy's, and Burchell's zebra, where we observed high numbers of satellite-free centromeres, karyotype reshuffing, and heterogeneity in satellite DNA families, CENP-B binding was rarely observed at primary constrictions. In the horse, where only one satellite-free centromere was found, most centromeres contain CENPB-sat and a subset of them binds CENP-B.

Taking together our results, we propose that the uncoupling between CENP-B and the centromeric core may drive the centromeric plasticity observed in equids. However, an important question remains open: despite being uncoupled to centromeres and poorly binding to DNA in some species, why is CENP-B well conserved and expressed in all equids? Is it simply the result of an evolutionary process or may CENP-B play extra-centromeric yet unknown roles?

Conclusions

In the mammalian species studied so far CENP-B and CENP-A bind the major centromeric satellite. Our study showed that, in equids, CENP-B was not detectable at the numerous satellite-free and at the majority of satellite-based centromeres while it was localized at several ancestral inactivated centromeres. CENP-B binding sites were also detected at intra-chromosomal loci suggesting that the protein may play extra-centromeric roles.

By comparing CENP-B positive and negative centromeres, which are naturally occurring in the equid system, we demonstrated that centromeres lacking CENP-B are functional and recruit normal amounts of the centromeric proteins CENP-A and CENP-C. Thus, differently from what previously shown in human and mouse experimental systems, we proved that, in the equid natural system, the role of CENP-B is more complex and that binding of CENP-A and CENP-C is not universally influenced by CENP-B. Although in cultured cells we did not observe any segregation defect of CENP-B negative centromeres, we cannot exclude but actually propose that, on an evolutionary time scale, and possibly in meiosis, minor perturbations in centromere function may favor the formation of neocentromeres through repositioning or Robertsonian fusion. Therefore, a low probability of mis-segregation during large numbers of cell divisions may eventually cause karyotype reshuffling favoring speciation.

The absence of CENP-B at most equid centromeres is related to the lack of CENP-B boxes rather than to peculiar features of the protein itself. While no CENP-B boxes were identified in the CENP-A binding domains of the satellite-free centromeres and in the major satellite repeat, this motif was found in a previously undescribed repeat. A comparative analysis of the localization of the CENP-B box containing satellite suggests that this satellite corresponds to an old centromeric repeat which was bound by CENP-A in the common ancestor of extant equid species. We propose that, during the radiation of *Equus* species, this satellite lost the centromeric function and the resulting uncoupling between CENP-B and CENP-A may have played a role in the evolutionary reshuffling of centromeres.

These findings open a new scenario for the study of the mysterious CENP-B protein, providing new insights into the complexity of centromere organization in a largely biodiverse world where the majority of mammalian species still have to be studied.

Methods

Cell lines

Primary fibroblast cell lines from horse, donkey, mule, Burchell's zebra and Grevy's zebra were previously described [39, 43, 46]. The fibroblast cell lines immortalized by telomerase were previously described [43, 57].

Fibroblasts were cultured in high-glucose DMEM medium, supplemented with 20% fetal bovine serum, 2 mM glutamine, 2% non-essential amino acids, and 1% penicillin/ streptomycin. HeLa cells were cultured in high-glucose DMEM medium, supplemented with 10% fetal bovine serum, 2 mM glutamine, 2% non-essential amino acids, and 1% penicillin/streptomycin. Cells were maintained in a humidified atmosphere of 5% CO_2 at 37 °C. All the cell lines tested negative for mycoplasma.

Antibodies

Four different commercial polyclonal anti-CENP-B antibodies were used: sc-22788 (Santa Cruz Biotechnology Inc.), raised against amino acids 535–599 mapping at the C-terminus of human CENP-B (P07199) (Additional file 1: Fig. S1); ab84489 (Abcam), raised against a synthetic peptide corresponding to the 540–599 residues of human CENPB (P07199) (Additional file1: Fig. S1); the H00001059-B01P (Abnova) and the 07–735 (Sigma-Aldrich) were obtained using the entire human CENP-B protein as immunogen. Preliminary immunofluorescence experiments were carried out on human HeLa cells and horse and donkey fibroblasts to compare the three antibodies sc-22788 (Santa Cruz Biotechnology Inc.), ab84489 (Abcam), and H00001059-B01P (Abnova). The results of this comparison are shown in Additional file 1: Fig. S6.

Anti-CENP-A and anti-CENP-C sera were previously described [79, 80].

ChIP-seq

Chromatin from primary fibroblasts 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 [43] using the anti-CENP-B sc-22788 antibody (Santa Cruz Biotechnology Inc.). Paired-end sequencing was performed with Illumina HiSeq2000 and Illumina HiSeq2500 platforms by IGA Technology Services (Udine, Italy). Reads from ChIP-seq experiments with the anti-CENP-A antibody were previously described [43, 46] and deposited in NCBI SRA Archive (SRR27325169, SRR27325168, SRR5515973, SRR5515972, SRR17956804, SRR17956803, SRR17956806, SRR17956805). The details of each dataset are reported in Additional file 1: Table S13.

Identification of the CENP-B bound satellite, CENPB-sat, from ChIP-seq data

Reads from the ChIP-seq experiment with the anti-CENP-B antibody on horse primary fibroblasts were aligned to the horse reference genome (EquCab 2.0, 2007 release) with Bowtie (version 1.1.2), using the single end mode and k = 10 correction in order to refine the mapping of reads from satellite repeats [81].

Peak calling was performed using MACS14 (version 1.4.1) [82]. Stringency criteria were: chrUn selection, fold enrichment > 8, $-10Log_{10}$ (p-Value) > 100 and FDR (%) < 1. The 57 top-ranked regions were analyzed through Tandem Repeat Finder [83]. For each region, Tandem Repeat Finder reports one or more classes of tandem repeats, providing a consensus for each class. The 425 bp consensus sequence of CENPB-sat was obtained by Multalin [84] alignment of sequences containing a canonical CENP-B box. Consensus sequences other than CENPB-sat identified by Tandem Repeat Finder were analyzed by RepeatMasker (Galaxy Version 4.1.5 + galaxy0) using the RepBase library (release October 26, 2018).

To evaluate enrichment and genomic abundance of CENPB-sat in the four species, ChIP and Input reads were mapped with Bowtie2.0 (2.4.2 version) [85] using the single end mode and default parameter on the consensus sequences of the entire horse CENPB-sat or the 201 bp fragment containing the CENP-B box and not showing any identity with 37cen and ERE-1 (SAT_EC and D26566 in RepBase). Counts per million (CPM) from resulting BAM files were obtained using idxstats command from the Samtools package (version 1.15.1) [86]. The consensus of the CENP-B box was deduced from the Input reads of each species aligned to the horse CENPB-sat sequence using the "Copy consensus sequence" function of the IGV software (2.9.2 version).

Identification of satellite repeats from unassembled Input reads was performed with TAREAN (Galaxy Version 2.3.8.1), a computational pipeline that uses graph-based repeat clustering to detect satellite repeats directly from unassembled short reads [55] using 2 million reads as sample size and default parameters which allow high confidence outputs. Since no satellite containing a CENP-B box was identified using Burchell's zebra Input reads, we run the same analysis using ChIP reads obtained using the anti-CENP-B antibody. ChIP-seq mapper (Galaxy Version 0.1.1) [87] was used to evaluate the enrichment of satellite repeats identified by TAREAN in ChIP-seq experiments performed with anti-CENP-A antibody [43, 46].

Detection of dyad symmetries and other non-B form DNA motifs

Dyad symmetries were searched in the centromeric regions we previously assembled [37, 43, 46] using EMBOSS Palindrome (version 6.6.0–7) with the minimum palindrome being 5, the maximum palindrome being 100, allowing a gap limit of 20 and allowing overlapping dyad symmetries as previously described [30, 56]. Non-B DNA-forming sequence motifs, including A-phased repeats, direct repeats, inverted repeats, mirror repeats, Z-DNA, and G-quadruplex, were predicted using Non-B DB v2.0 [88]. For each sequence of interest, we computed the number and the coverage of sequences forming a dyad or other non-B motifs and normalized per kilobase. Centromeres containing DNA duplications [43, 46] were excluded from this analysis.

For each species, we randomly selected 100 control genomic region with a similar GC content $(37 \pm 1.5 \text{ for the horse}, 35 \pm 1.5 \text{ for the donkey}, 36.6 \pm 1.5 \text{ for the Grevy's zebra}, and 37 \pm 1.5 \text{ for the Burchell's zebra}; these ranges correspond to the average GC \pm the standard deviation of the GC content of the centromeric regions) and a length corresponding to the average length of the centromeric domains (500 kb for the horse, 404 kb for the donkey, 237 kb for the Grevy's zebra, and 225 kb for the Burchell's zebra). The selection of control regions was performed using bedtools (v2.30.0) and Seqkit (v2.6.1). Reference genomes used to identify control regions were the horse EquCab2.0 assembly [37], the donkey ASM1607732v2 assembly, the Grevy's zebra Equus_grevyi_HiC assembly [89], and the Burchell's zebra Equus_quagga_HiC assembly [89].$

We calculated standardized Z-score for the values of the unique horse satellite-free centromere (ECA11) with respect to control regions. To test whether the differences were statistically significant, we calculated the *P*-values using Z-score calculator [90]. For the other species, in case of normal distribution, unpaired, two-tailed *t*-test or unpaired two-tailed Welch's *t*-test were used [91]. In the case of non-Gaussian distribution, two-tailed Mann–Whitney *U* test was applied [90]. Boxplots with statistical significance analysis were obtained using ggplot2 and ggsignif R packages.

Genome-wide analysis of CENP-B binding sites

To evaluate the CENP-B binding profiles at satellite-free CENP-A binding domains, ChIP-seq reads were aligned with Bowtie2 (version 2.4.2) using paired-end mode and

default parameters to the species-specific references: EquCab2.0 for the horse [37], Equ-CabAsiB [43], EBU_EGR_cen [46] for the Grevy's zebra, and Equus_quagga_cen [46] for the Burchell's zebra. Low-quality aligments (MAPQ < 20) were filtered out, and normalized enrichment peaks were obtained with the bamCompare tool available in the deep-Tools suite (3.5.0 version) [92] using RPKM normalization in subtractive mode. Plots were obtained with pyGenomeTracks (3.6 version) [93].

To evaluate the genome-wide distribution of CENP-B binding sites, ChIP-seq reads from the four species were aligned with paired-end mode to the EquCab2.0 reference genome with Bowtie2 (version 2.4.2) using default parameters [81, 85]. Peak calling was performed with MACS2 (version 2.2.7.1) [82] using 0.01 as q-value cutoff. We excluded from the analysis the peaks overlapping satellite sequences using UCSC Table Browser and the peaks identified in unplaced contigs. CENP-B boxes were searched using FIMO [94]. The content of interspersed repeats was analyzed with RepeatMasker using the RepBase library (release October 26, 2018). Bedtools (v2.30.0) was utilized to identify peaks shared among different species.

Sequencing of CENP-B genes

The sequence of the *CENP-B* coding sequence of horse, donkey, Grevy's zebra, and Burchell's zebra was obtained by Sanger sequencing of PCR fragments and by directly assembling reads from ChIP-seq input datasets. Primers used for PCR amplification and sequencing are listed in Additional file 1: Table S14.

Western blotting

Total protein extracts were prepared from samples of three million cells as follows: the cells were washed twice with ice cold 1xPBS, resuspended in lysis buffer (50 mM Tris–HCl pH 6.8, 86 mM β -mercaptoethanol, 2% SDS) and boiled for 10 min, as previously described [54]. Nuclear and cytoplasmic protein extracts were prepared using the fractionation protocol developed by Suzuki and colleagues [95]. Briefly, starting from samples of 30 million cells, the cells were resuspended in ice-cold 0.1% NP40 in PBS. The nuclear and cytoplasmic fractions were then separated by a 10-s centrifugation at 6000 rpm. The supernatant was saved as cytoplasmic fraction, diluted in Laemmli buffer and boiled for 1 min. The pellet was resuspended in ice-cold 0.1% NP40 in PBS, centrifuged again as above and the final pellet was resuspended in Laemmli buffer, sonicated, boiled for 1 min, and saved as nuclear extract.

Proteins were separated by SDS-PAGE on polyacrilamide gel and blotted to nitrocellulose membranes (AmershamTM HybondTM-ECL, GE-Healthcare) according to standard methods. Membranes were incubated with the anti- α tubulin antibody [DM1A] ab7291 (Abcam), diluted 1:5000, the anti-CENP-B sc-22788 antibody (Santa Cruz Biotechnology Inc.), diluted 1:750 or with the anti-CENP-B 07–735 antibody (Sigma-Aldrich), and diluted 1:1000. HRP conjugated secondary antibodies were used. Pre-incubation of membranes and dilutions of antibodies were performed in 1 × PBS containing 0.05% Tween-20 and 7.5% skim milk. Detection was performed using the BioRad ClarityTM Western ECL Substrate kit following manufacturer's procedures.

CENPB-sat plasmid vector construction

The portion of the CENPB-sat comprising the CENP-B box and lacking identity regions with the 37cen satellite was amplified from horse genomic DNA using the following primer oligonucleotides containing EcoRI and SalI adapters required for cloning purposes: CENPBsat-F 5'-ATTGAATTCCCTTTCTGACATAGGTGCTTT CTG-3' and CENPBsat-R 5'- ATTGTCGACGCTTTAGGACTTCTGGTTCTG-3'. PCR products were digested with EcoRI/SalI and cloned in the pSVal plasmid [96]. An 8-copies array of the cloned portion was obtained as previously described [53].

Immunofluorescence and FISH

We carried out preliminary immunofluorescence experiments to test several permeabilization and fixation procedures with the three anti-CENP-B antibodies described above (sc-22788 Santa Cruz Biotechnology Inc., ab84489 Abcam or H00001059-B01P Abnova). The best combination was fixation with ice-cold methanol for 4 min followed by permeabilization with $1 \times PBS 0.05\%$ Tween-20 for 15 min at room temperature and incubation at 37 °C for 2 h with H00001059-B01P Abnova antibody diluted 1:100. Incubation with the anti-CENP-A [80] or anti-CENP-C serum [79], both diluted 1:100, was carried out at 37 °C for 1 h. Digital grey-scale images were acquired with a fluorescence microscope (Zeiss Axio Scope.A1) equipped with a cooled CCD camera (Photometrics) using a $63 \times oil$ objective. In immuno-FISH experiments, immunofluorescence signals were collected before hybridization with the CENPB-sat FISH probe. Pseudo-coloring and merging of images were performed using the IpLab software.

Metaphase spreads were obtained with the standard air-drying procedure. CENPBsat plasmid extraction, nick translation with Cy3-dUTP (ENZ-42501), and hybridization were performed as previously described [39]. Chromosomes were counterstained with DAPI and identified by computer-generated reverse DAPI banding according to the published karyotypes.

3D-immunofluorescence on whole cells was performed using a slight modification of the protocol described by Solovei and Cremer [97]. Cells were grown on coverslips, rinsed with PBS, and fixed with 4% paraformaldehyde in PBS at room temperature. During the last minute of fixation, a few drops of 1 × PBS 0.5% Triton X-100/PBS were added. After three washes in 0.01% Tween-20, cells were permeabilized with 1 × PBS 0.5% Tween-20 for 20 min at room temperature. Anti-CENP-B (sc-22788 Santa Cruz Biotechnology Inc.) and anti-tubulin (ab7291 Abcam) antibodies were diluted 1:80 and 1:500, respectively. Stacks of optical sections through whole cells were collected using a Leica TCS SP8 STED 3X confocal microscope (Centro Grandi Strumenti, University of Pavia).

Quantification of immunofluorescence signals

Quantification of CENP-A, CENP-B, and CENP-C signal intensities on metaphase spreads was performed using the Fiji software [98]. The integrated signal density of each centromeric signal was calculated by subtracting the fluorescence intensity of the back-ground from the total intensity of the signal. Statistical significance was evaluated using Spearman's Rho correlation test and unpaired two-tailed *t*-test [90]. Boxplots with statistical significance analysis were obtained using ggplot2 and ggsignif R packages.

Statistical analysis of the fraction of CENP-B positive centromeres in immortalized cell lines was performed by a one-way ANOVA followed by Tukey's post hoc test and Spearman's correlation test.

CENP-B-eGFP plasmid construction and transfection

The horse *CENP-B* coding sequence was cloned upstream of the enhanced Green Fluorescent Protein (eGFP) cDNA, into an expression vector that was previously constructed in our laboratory [96]. The vector contains the puromycin resistance genes. The chimeric protein was expressed under the control of the Cytomegalovirus-immediate early (CMVie) promoter.

The plasmid (pCCB-GFP) was used to transfect a horse fibroblast cell line, previously immortalized in our laboratory [57]. Transfection was carried out using the NeonTM Transfection System (Thermo Fisher Scientific) according to the manufacturer's protocol. Forty-eight h after transfection, puromycin (750 ng/ml) was added, and resistant clones were isolated after 3 weeks. Cells were then harvested by trypsinization, treated with hypothonic 75 mM KCl solution for 25 min at 37 °C, cyto-spun onto slides at 1250 rpm for 8 min, and fixed with ice-cold methanol for 4 min.

Interphase aneuploidy analysis

For each experiment, 3×10^5 cells were seeded in 10 cm plates. Untreated cells were grown for 72 h, while treated cells were exposed to 200 nM nocodazole (Sigma-Aldrich) after 24 h culture period and then grown for the remaining 48 h. Cells were then harvested by trypsinization, treated with hypothonic solution (75 mM KCl) for 25 min at 37 °C, and then fixed with cold 1:3 acetic acid to methanol solution overnight at 4 °C. Nuclei were then spread onto slides according to the standard air-drying procedure.

To identify horse chromosomes 9 and 10, two bacterial artificial chromosomes derived from the CHORI-241 BAC library (CH241-361E21, chr9:36,816,509–36,983,616 in Equ-Cab3.0; CH241-403K5, chr10:28,469,205–28,662,312 in EquCab3.0) were extracted from 10 ml bacterial cultures with the Quantum Prep Plasmid miniprep kit (BioRad), according to supplier instructions. The probes were labeled by nick translation with Cy3-dUTP (Enzo Life Sciences), and FISH was performed as previously described [39]. The χ^2 test was used to evaluate whether the differences in the frequency of an euploid nuclei were statistically significant.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-025-03490-0.

Additional file 1: This file includes Supplementary Text, Supplementary Figs. S1 to S9, Supplementary Tables S1, S3, S4, S5, S8, S10, S11, S12, S13 and S14.

Additional file 2. This file includes Supplementary Tables S2, S6, S7 and S9.

Additional file 3: Movie S1 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in HeLa cells. Arrows indicate the cellsshown in Fig. 4.

Additional file 4: Movie S2 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in horse primary fibroblasts. Arrows indicate the cells shown in Fig. 4.

Additional file 5: Movie S3 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in horse primary fibroblasts. The arrow indicates the cell shown in Fig. 4.

Additional file 6: Movie S4 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in Grevy's zebra primary fibroblasts. Arrows indicate the cells shown in Fig. 4.

Additional file 7: Movie S5 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in donkey primary fibroblasts. Arrows indicate the cells shown in Fig. 4.

Additional file 8: Movie S6 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in donkey primary fibroblasts. Arrows indicate the cells shown in Fig. 4.

Additional file 9: Movie S7 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in Burchell's zebra primary fibroblasts. Arrows indicate the cells shown in Fig. 4.

Additional file 10: Movie S8 (related to Fig. 4). 3D fluorescence imaging of CENP-B (red) and tubulin (green) in Burchell's zebra primary fibroblasts. Arrows indicate the cells shown in Fig. 4.

Additional file 11: Review history.

Acknowledgements

We would like to thank Terje Raudsepp (Texas A&M University, USA) for providing to us the horse BAC clones, Douglas F. Antczak and Donald Miller (Cornell University, USA) for the mule fibroblast cell line, Sergio Comincini (University of Pavia) for helpful suggestions on cell transfection by electroporation, Anna Garbelli (Istituto di Genetica Molecolare, IGM-CNR, Pavia) for technical support in Chemidoc imaging system for western blotting experiments, Patrizia Vaghi and Amanda Oldani of Centro Grandi Strumenti - Confocal Microscopy Facility (University of Pavia) for their support and assistance in confocal imaging, and Giulio Pavesi (University of Milan), Riccardo Gamba, and Kevin Sullivan (University of Galway, Ireland) for advice on bioinformatic analyses and plasmid construction.

Peer review information

Andrew Cosgrove was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Review history

The review history is available as Additional file 11.

Authors' contributions

EC and FMP carried out most molecular and cell biology experiments and bioinformatic analyses. SGN, MB, EG, and ER contributed to some molecular and cell biology experiments. EG, EC, and FMP conceived the study and wrote the manuscript. EG supervised the study. All authors participated in discussions and result interpretation. All authors read and approved the final manuscript.

Funding

This research was funded by Animal Breeding and Functional Annotation of Genomes (A1201) Grant 2019–67015-29340/ Project Accession 1018854 from the USDA National Institute of Food and Agriculture, Italian Ministry of Education, University and Research (MIUR) (Dipartimenti di Eccellenza Program (2018–2022)—Department of Biology and Biotechnology "L. Spallanzani," University of Pavia).

The Galaxy server that was used for some calculations is in part funded by Collaborative Research Centre 992 Medical Epigenetics (DFG grant SFB 992/1 2012) and German Federal Ministry of Education and Research (BMBF grants 031 A538A/A538C RBC, 031L0101B/031L0101C de.NBI-epi, 031L0106 de.STAIR (de.NBI)). Computational resources for Repeat-Explorer analysis were provided by the ELIXIR-CZ project (LM2023055), part of the international ELIXIR infrastructure.

Data availability

Raw sequencing data from this study are available in the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA1054998 [99]. In this work, we also used publicly available ChIP-seq datasets (SRR27325169 [100], SRR27325168 [101], SRR5515973 [102], SRR5515972 [103], SRR17956804 [104], SRR17956803 [105], SRR17956806 [106], SRR17956805 [107]) that we previously deposited in NCBI SRA Archive. Microscopy images are available on Figshare (https://doi.org/10.6084/m9.figshare.28243004 [108]).

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 23 April 2024 Accepted: 29 January 2025

Published online: 06 February 2025

References

- 1. Kalitsis P, Choo KH. The evolutionary life cycle of the resilient centromere. Chromosoma. 2012;121(4):327-40.
- 2. Choo KH. Centromerization. Trends Cell Biol. 2000;10(5):182-8.
- 3. Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. Science. 2001;293(5532):1098–102.
- Cleveland DW, Mao Y, Sullivan KF. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell. 2003;112(4):407–21.
- 5. Plohl M, Luchetti A, Mestrović N, Mantovani B. Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. Gene. 2008;409(1–2):72–82.
- Fry K, Salser W. Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat Dipodomys ordii and characterization of similar sequences in other rodents. Cell. 1977;12(4):1069–84.
- 7. Garrido-Ramos MA. Satellite DNA: An Evolving Topic. Genes (Basel). 2017;8(9):230.
- Shepelev VA, Alexandrov AA, Yurov YB, Alexandrov IA. The evolutionary origin of man can be traced in the layers of defunct ancestral alpha satellites flanking the active centromeres of human chromosomes. PLoS Genet. 2009;5(9): e1000641.
- Kursel LE, Malik HS. The cellular mechanisms and consequences of centromere drive. Curr Opin Cell Biol. 2018;52:58–65.
- Allshire RC, Karpen GH. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet. 2008;9(12):923–37.
- 11. Sullivan KF, Hechenberger M, Masri K. Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. J Cell Biol. 1994;127(3):581–92.
- 12. Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. J Cell Biol. 1989;109(5):1963–73.
- 13. Kipling D, Warburton PE. Centromeres, CENP-B and Tigger too. Trends Genet. 1997;13(4):141–5.
- 14. Gamba R, Fachinetti D. From evolution to function: two sides of the same CENP-B coin? Exp Cell Res. 2020;390(2): 111959.
- Tanaka Y, Nureki O, Kurumizaka H, Fukai S, Kawaguchi S, Ikuta M, et al. Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. EMBO J. 2001;20(23):6612–8.
- 16. Tawaramoto MS, Park SY, Tanaka Y, Nureki O, Kurumizaka H, Yokoyama S. Crystal structure of the human centromere protein B (CENP-B) dimerization domain at 1.65-A resolution. J Biol Chem. 2003;278(51):51454–61.
- Logsdon GA, Gambogi CW, Liskovykh MA, Barrey EJ, Larionov V, Miga KH, et al. Human artificial chromosomes that bypass centromeric DNA. Cell. 2019;178(3):624-39.e19.
- 18. Kapoor M, Montes Oca Luna R de, Liu G, Lozano G, Cummings C, Mancini M, et al. The cenpB gene is not essential in mice. Chromosoma. 1998;107(8):570–6.
- 19. Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, Trowell H, et al. Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. J Cell Biol. 1998;141(2):309–19.
- 20. Fowler KJ, Hudson DF, Salamonsen LA, Edmondson SR, Earle E, Sibson MC, et al. Uterine dysfunction and genetic modifiers in centromere protein B-deficient mice. Genome Res. 2000;10(1):30–41.
- 21. Dai X, Otake K, You C, Cai Q, Wang Z, Masumoto H, et al. Identification of novel α-n-methylation of CENP-B that regulates its binding to the centromeric DNA. J Proteome Res. 2013;12(9):4167–75.
- 22. Fachinetti D, Han JS, McMahon MA, Ly P, Abdullah A, Wong AJ, et al. DNA sequence-specific binding of CENP-B enhances the fidelity of human centromere function. Dev Cell. 2015;33(3):314–27.
- Dumont M, Gamba R, Gestraud P, Klaasen S, Worrall JT, De Vries SG, et al. Human chromosome-specific aneuploidy is influenced by DNA-dependent centromeric features. Embo j. 2020;39(2): e102924.
- 24. Abdel-Hafiz HA, Schafer JM, Chen X, Xiao T, Gauntner TD, Li Z, et al. Y chromosome loss in cancer drives growth by evasion of adaptive immunity. Nature. 2023;619(7970):624–31.
- McNulty SM, Sullivan LL, Sullivan BA. Human centromeres produce chromosome-specific and array-specific alpha satellite transcripts that are complexed with CENP-A and CENP-C. Dev Cell. 2017;42(3):226-40.e6.
- Morozov VM, Giovinazzi S, Ishov AM. CENP-B protects centromere chromatin integrity by facilitating histone deposition via the H3.3-specific chaperone Daxx. Epigenetics Chromatin. 2017;10(1):63.
- 27. Kumon T, Ma J, Akins RB, Stefanik D, Nordgren CE, Kim J, et al. Parallel pathways for recruiting effector proteins determine centromere drive and suppression. Cell. 2021;184(19):4904-18.e11.
- Cam HP, Noma K, Ebina H, Levin HL, Grewal SI. Host genome surveillance for retrotransposons by transposonderived proteins. Nature. 2008;451(7177):431–6.

- Zaratiegui M, Vaughn MW, Irvine DV, Goto D, Watt S, Bähler J, et al. CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. Nature. 2011;469(7328):112–5.
- 30. Kasinathan S, Henikoff S. Non-B-form DNA is enriched at centromeres. Mol Biol Evol. 2018;35(4):949-62.
- 31. Nagpal H, Ali-Ahmad A, Hirano Y, Cai W, Halic M, Fukagawa T, et al. CENP-A and CENP-B collaborate to create an open centromeric chromatin state. Nat Commun. 2023;14(1):8227.
- 32. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature. 2013;499(7456):74–8.
- Trifonov VA, Musilova P, Kulemsina AI. Chromosome evolution in Perissodactyla. Cytogenet Genome Res. 2012;137(2–4):208–17.
- 34. Trifonov VA, Stanyon R, Nesterenko AI, Fu B, Perelman PL, O'Brien PC, et al. Multidirectional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla. Chromosome Res. 2008;16(1):89–107.
- 35. Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, et al. Speciation with gene flow in equids despite extensive chromosomal plasticity. Proc Natl Acad Sci U S A. 2014;111(52):18655–60.
- 36. Carbone L, Nergadze SG, Magnani E, Misceo D, Francesca Cardone M, Roberto R, et al. Evolutionary movement of centromeres in horse, donkey, and zebra. Genomics. 2006;87(6):777–82.
- Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, et al. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science. 2009;326(5954):865–7.
- 38. Piras FM, Nergadze SG, Poletto V, Cerutti F, Ryder OA, Leeb T, et al. Phylogeny of horse chromosome 5q in the genus Equus and centromere repositioning. Cytogenet Genome Res. 2009;126(1–2):165–72.
- Piras FM, Nergadze SG, Magnani E, Bertoni L, Attolini C, Khoriauli L, et al. Uncoupling of satellite DNA and centromeric function in the genus Equus. PLoS Genet. 2010;6(2): e1000845.
- Musilova P, Kubickova S, Vahala J, Rubes J. Subchromosomal karyotype evolution in Equidae. Chromosome Res. 2013;21(2):175–87.
- Purgato S, Belloni E, Piras FM, Zoli M, Badiale C, Cerutti F, et al. Centromere sliding on a mammalian chromosome. Chromosoma. 2015;124(2):277–87.
- 42. Giulotto E, Raimondi E, Sullivan KF. The unique DNA sequences underlying equine centromeres. Prog Mol Subcell Biol. 2017;56:337–54.
- Nergadze SG, Piras FM, Gamba R, Corbo M, Cerutti F, McCarter JGW, et al. Birth, evolution, and transmission of satellite-free mammalian centromeric domains. Genome Res. 2018;28(6):789–99.
- 44. Peng S, Petersen JL, Bellone RR, Kalbfleisch T, Kingsley NB, Barber AM, et al. Decoding the equine genome: lessons from ENCODE. Genes (Basel). 2021;12(11):1707.
- 45. Piras FM, Cappelletti E, Santagostino M, Nergadze SG, Giulotto E, Raimondi E. Molecular dynamics and evolution of centromeres in the genus Equus. Int J Mol Sci. 2022;23(8):4183.
- 46. Cappelletti E, Piras FM, Sola L, Santagostino M, Abdelgadir WA, Raimondi E, et al. Robertsonian fusion and centromere repositioning contributed to the formation of satellite-free centromeres during the evolution of zebras. Mol Biol Evol. 2022;39(8):msac162.
- 47. Piras FM, Cappelletti E, Abdelgadir WA, Salamon G, Vignati S, Santagostino M, et al. A satellite-free centromere in Equus przewalskii Chromosome 10. Int J Mol Sci. 2023;24(4):4134.
- 48. Cappelletti E, Piras FM, Sola L, Santagostino M, Petersen JL, Bellone RR, et al. The localization of centromere protein A is conserved among tissues. Commun Biol. 2023;6(1):963.
- 49. Cerutti F, Gamba R, Mazzagatti A, Piras FM, Cappelletti E, Belloni E, et al. The major horse satellite DNA family is associated with centromere competence. Mol Cytogenet. 2016;9:35.
- 50. Rocchi M, Archidiacono N, Schempp W, Capozzi O, Stanyon R. Centromere repositioning in mammals. Heredity (Edinb). 2012;108(1):59–67.
- 51. Kalbfleisch TS, Rice ES, DePriest MS, Walenz BP, Hestand MS, Vermeesch JR, et al. Improved reference genome for the domestic horse increases assembly contiguity and composition. Commun Biol. 2018;1:197.
- 52. DNA Zoo Website [cited 2023 31 August]. Available from: https://www.dnazoo.org/.
- 53. Nergadze SG, Belloni E, Piras FM, Khoriauli L, Mazzagatti A, Vella F, et al. Discovery and comparative analysis of a novel satellite, EC137, in horses and other equids. Cytogenet Genome Res. 2014;144(2):114–23.
- Santagostino M, Khoriauli L, Gamba R, Bonuglia M, Klipstein O, Piras FM, et al. Genome-wide evolutionary and functional analysis of the Equine Repetitive Element 1: an insertion in the myostatin promoter affects gene expression. BMC Genet. 2015;16:126.
- Novák P, Ávila Robledillo L, Koblížková A, Vrbová I, Neumann P, Macas J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res. 2017;45(12): e111.
- Patchigolla VSP, Mellone BG. Enrichment of non-B-form DNA at D. melanogaster centromeres. Genome Biol Evol. 2022;14(5):evac054.
- Vidale P, Magnani E, Nergadze SG, Santagostino M, Cristofari G, Smirnova A, et al. The catalytic and the RNA subunits of human telomerase are required to immortalize equid primary fibroblasts. Chromosoma. 2012;121(5):475–88.
- Vidale P, Piras FM, Nergadze SG, Bertoni L, Verini-Supplizi A, Adelson D, et al. Chromosomal assignment of six genes (EIF4G3, HSP90, RBBP6, IL8, TERT, and TERC) in four species of the genus Equus. Anim Biotechnol. 2011;22(3):119–23.
- 59. Zongaro S, de Stanchina E, Colombo T, D'Incalci M, Giulotto E, Mondello C. Stepwise neoplastic transformation of a telomerase immortalized fibroblast cell line. Cancer Res. 2005;65(24):11411–8.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, et al. Recent segmental duplications in the human genome. Science. 2002;297(5583):1003–7.
- Ventura M, Weigl S, Carbone L, Cardone MF, Misceo D, Teti M, et al. Recurrent sites for new centromere seeding. Genome Res. 2004;14(9):1696–703.
- 62. Talbert PB, Henikoff S. What makes a centromere? Exp Cell Res. 2020;389(2): 111895.
- 63. Suntronpong A, Kugou K, Masumoto H, Srikulnath K, Ohshima K, Hirai H, et al. CENP-B box, a nucleotide motif involved in centromere formation, occurs in a New World monkey. Biol Lett. 2016;12(3):20150817.

- 64. Kugou K, Hirai H, Masumoto H, Koga A. Formation of functional CENP-B boxes at diverse locations in repeat units of centromeric DNA in New World monkeys. Sci Rep. 2016;6:27833.
- 65. Yoda K, Nakamura T, Masumoto H, Suzuki N, Kitagawa K, Nakano M, et al. Centromere protein B of African green monkey cells: gene structure, cellular expression, and centromeric localization. Mol Cell Biol. 1996;16(9):5169–77.
- Goldberg IG, Sawhney H, Pluta AF, Warburton PE, Earnshaw WC. Surprising deficiency of CENP-B binding sites in African green monkey alpha-satellite DNA: implications for CENP-B function at centromeres. Mol Cell Biol. 1996;16(9):5156–68.
- 67. Roberti A, Bensi M, Mazzagatti A, Piras FM, Nergadze SG, Giulotto E, et al. Satellite DNA at the centromere is dispensable for segregation fidelity. Genes (Basel). 2019;10(6):469.
- 68. She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, et al. The structure and evolution of centromeric transition regions within the human genome. Nature. 2004;430(7002):857–64.
- 69. Cacheux L, Ponger L, Gerbault-Seureau M, Loll F, Gey D, Richard FA, et al. The targeted sequencing of alpha satellite DNA in Cercopithecus pogonias provides new insight into the diversity and dynamics of centromeric repeats in Old World monkeys. Genome Biol Evol. 2018;10(7):1837–51.
- 70. Repenning CA, Weasma TR, Scott GR. The early Pleistocene (latest Blancan-earliest Irvingtonian) Froman Ferry fauna and history of the Glenns Ferry Formation, southwestern Idaho. Report. 1995. Report No.: 2105.
- Bernor RL, Cirilli O, Jukar AM, Potts R, Buskianidze M, Rook L. Evolution of early Equus in Italy, Georgia, the Indian Subcontinent, East Africa, and the origins of African zebras. Front Ecol Evol. 2019;7(166). https://doi.org/10.3389/ fevo.2019.00166.
- Marshall OJ, Chueh AC, Wong LH, Choo KH. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum Genet. 2008;82(2):261–82.
- Tanaka Y, Kurumizaka H, Yokoyama S. CpG methylation of the CENP-B box reduces human CENP-B binding. FEBS J. 2005;272(1):282–9.
- 74. Altemose N, Logsdon GA, Bzikadze AV, Sidhwani P, Langley SA, Caldas GV, et al. Complete genomic and epigenetic maps of human centromeres. Science. 2022;376(6588):eabl4178.
- Heslop-Harrison JS, Schwarzacher T. Nucleosomes and centromeric DNA packaging. Proc Natl Acad Sci U S A. 2013;110(50):19974–5.
- 76. Hasson D, Panchenko T, Salimian KJ, Salman MU, Sekulic N, Alonso A, et al. The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat Struct Mol Biol. 2013;20(6):687–95.
- 77. Talbert PB, Henikoff S. The genetics and epigenetics of satellite centromeres. Genome Res. 2022;32(4):608–15.
- Mohibi S, Srivastava S, Wang-France J, Mirza S, Zhao X, Band H, et al. Alteration/deficiency in activation 3 (ADA3) protein, a cell cycle regulator, associates with the centromere through CENP-B and regulates chromosome segregation. J Biol Chem. 2015;290(47):28299–310.
- 79. Trazzi S, Perini G, Bernardoni R, Zoli M, Reese JC, Musacchio A, et al. The C-terminal domain of CENP-C displays multiple and critical functions for mammalian centromere formation. PLoS ONE. 2009;4(6): e5832.
- Cappelletti E, Piras FM, Badiale C, Bambi M, Santagostino M, Vara C, et al. CENP-A binding domains and recombination patterns in horse spermatocytes. Sci Rep. 2019;9(1):15800.
- 81. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
- 82. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012;7(9):1728-40.
- 83. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573-80.
- 84. Corpet F. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 1988;16(22):10881–90.
- 85. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078–9.
- 87. Neumann P, Navrátilová A, Schroeder-Reiter E, Koblížková A, Steinbauerová V, Chocholová E, et al. Stretching the rules: monocentric chromosomes with multiple centromere domains. PLoS Genet. 2012;8(6): e1002777.
- 88. Cer RZ, Donohue DE, Mudunuri US, Temiz NA, Loss MA, Starner NJ, et al. Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. Nucleic Acids Res. 2013;41(Database issue):D94-D100.
- 89. DNA Zoo Website [Available from: https://www.dnazoo.org/.
- 90. Social Science Statistics Website [Available from: https://www.socscistatistics.com/.
- 91. VassarStats: Website for Statistical Computation [Available from: http://vassarstats.net/.
- 92. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160–5.
- Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics. 2021;37(3):422–3.
- 94. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017–8.
- 95. Suzuki K, Bose P, Leong-Quong RY, Fujita DJ, Riabowol K. REAP: A two minute cell fractionation method. BMC Res Notes. 2010;3:294.
- 96. Nergadze SG, Farnung BO, Wischnewski H, Khoriauli L, Vitelli V, Chawla R, et al. CpG-island promoters drive transcription of human telomeres. RNA. 2009;15(12):2186–94.
- Solovei I, Cremer M. 3D-FISH on cultured cells combined with immunostaining. Methods Mol Biol. 2010;659:117–26.
- Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676–82.
- Giulotto E. ChIP-seq with anti-CENP-B antibody on chromatin extracted from primary fibroblasts of horse, donkey, Grevy's zebra and Burchell's zebra. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/ bioproject/?term=PRJNA1054998 (2023).
- 100. University of Pavia. Sequencing of horse centromeres. HSF fibroblasts ChIP. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/?term=SRR27325169.

- 101. University of Pavia. Sequencing of horse centromeres. HSF fibroblasts Input. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/?term=SRR27325168.
- University of Pavia. Horse and donkey centromeres. DonkeyB-rep1-IP. Sequence Read Archive. https://www.ncbi. nlm.nih.gov/sra/?term=SRR5515973.
- 103. University of Pavia. Horse and donkey centromeres. DonkeyB-rep1.2-Input. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/?term=SRR5515972.
- 104. University of Pavia. Centromeres of Burchell's and Grevy's zebras. EGR CENP-A ChIP. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/?term=SRR17956804.
- 105. University of Pavia. Centromeres of Burchell's and Grevy's zebras. EGR Input. Sequence Read Archive. https://www. ncbi.nlm.nih.gov/sra/?term=SRR17956803.
- University of Pavia. Centromeres of Burchell's and Grevy's zebras. EBU CENP-A ChIP. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/?term=SRR17956806.
- University of Pavia. Centromeres of Burchell's and Grevy's zebras. EBU Input. Sequence Read Archive. https://www. ncbi.nlm.nih.gov/sra/?term=SRR17956805.
- Cappelletti E, Piras FM, Biundo M, Raimondi E, Nergadze S, Giulotto E. CENP-A/CENP-B uncoupling in the evolutionary reshuffling of centromeres in equids. Figshare. https://doi.org/10.6084/m9.figshare.28243004.v1 (2025).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.