UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie
"Lazzaro Spallanzani"

Laurea Magistralis in Molecular Biology and Genetics

Organization of centromeres in two Artiodactyla species: *Nanger dama* and *Nanger soemmerringii*

Supervisor:
Prof.ssa Elena Giulotto

Co-supervisors:
Dott.ssa Eleonora Cappelletti
Dott.ssa Maria Laura Biundo

Experimental thesis by
Yary Ferretti

Academic Year 2024/2025

TABLE OF CONTENTS

ABSTRACT

The centromere is a specialized nucleoprotein structure of the eukaryotic chromosome that ensures the accurate segregation of sister chromatids during mitosis and meiosis. By serving as the assembly site for the kinetochore, it guarantees the maintenance of genomic integrity during cell divisions. Despite its essential and highly conserved function, the DNA sequences underlying centromeric loci exhibit remarkable variability, even between closely related species. This phenomenon, known as the "centromere paradox," is explained by the fact that centromere identity is not determined by a specific DNA sequence but is instead epigenetically defined by the binding of the centromere-specific histone H3 variant, CENP-A.

In most mammalian species, centromeres are characterized by large, tandemly repeated arrays of satellite DNA. While these sequences are thought to contribute to centromere stability and heterochromatin organization, they are not strictly required for centromeric function. The discovery of "neocentromeres", functionally active centromeres that arise at novel chromosomal loci devoid of satellite DNA, has provided profound insights into the epigenetic nature of centromeres. Our laboratory has previously identified several satellite-free neocentromeres fixed within the genus *Equus*, that underwent rapid and recent evolution, establishing it as an exceptional model system for studying chromosomal evolution and centromere maturation. Recent investigations have extended these findings to other Perissodactyla, such as *Tapirus indicus*, whose karyotype was reshaped by evolutionarily recent changes, suggesting that satellite-free centromeres may be more common than previously assumed.

In this thesis work, we aimed to investigate whether evolutionary recent satellite-free centromeres are also present in the order Artiodactyla. We focused on two gazelle species within the genus *Nanger*: *N. dama* and *N. soemmerringii*, because of their high genomic instability and plastic karyotypes, similarly to equid species and *T. indicus*. The karyotypes of gazelles are characterized by variable diploid numbers (2n = 30–58) resulting from various chromosomal rearrangements, predominantly Robertsonian fusions.

To characterize the centromeric landscape of *N. dama* and *N. soemmerringii*, we employed a multidisciplinary approach combining cytogenetic and molecular techniques. We performed Fluorescence *In Situ* Hybridization (FISH) experiments to visualize the distribution of satellite DNA on metaphase chromosomes. To precisely map the functional centromeric domains at the DNA sequence level, we performed Chromatin Immunoprecipitation followed by sequencing (ChIP-seq) using a CREST serum containing antibodies against centromeric proteins, including CENP-A. We leveraged a preliminary draft genome assembly provided by the Ruminant Telomere-to-Telomere

(RT2T) Consortium as a reference for mapping. Moreover, we performed bioinformatic analyses with TAREAN for satellite DNA identification and BLAST to investigate telomeric repeats distribution.

TAREAN analysis detected two distinct satellite DNA families: NangerSat1 (701 bp long) and NangerSat2 (785 bp long). NangerSat1 was identified as the major centromeric satellite in both *Nanger* species, sharing high identity with other Artiodactyl satellites like OSSAT1 and BTSAT6. NangerSat2, while conserved with other known satellites among Artiodactyls, was found to be pericentromeric rather than centromeric in *N. dama*.

Cytogenetic FISH experiments with NangerSat1 used as a probe revealed that satellite DNA signals are present in the majority of primary constrictions in both species. Coversely, three chromosomes in each species (two submetacentric and one acrocentric) lacked detectable satellite signals, providing the first cytogenetic evidence of potential centromeres devoid of satellite DNA in these gazelles.

Molecular characterization by ChIP-seq approach confirmed the coexistence of satellite-based and satellite-free centromeres. In *N. dama*, we identified 13 canonical satellite-based centromeres and 3 satellite-free centromeres. Similarly, in *N. soemmerringii*, we identified 14 satellite-based centromeres and 3 satellite-free centromeres. The satellite-free domains in both species are located on chromosomal sequences previously designated as XY1, XY2, and XY4 in the RT2T draft assembly. Genomic comparative analysis between *N. dama* draft assembly and cattle (*Bos taurus*) genome, allowed us to assign these uncharacterized sequences to specific chromosomes, contributing to the refinement of the genome assembly. Specifically, the chromosomes named X1, Y1 and X2 carry satellite-free centromeres.

A significant finding of this work was the identification of heterochromatic Interstitial Telomeric Sequences (het-ITSs) in the pericentromeric domains of *N. dama*. These large clusters of telomeric-like repeats (TTAGGG) were identified both cytogenetically and molecularly. Our analysis suggests that these het-ITSs represent molecular signatures of ancestral chromosomal fusions.

The results of this study demonstrate that the occurrence of satellite-free centromeres is not limited to the order Perissodactyla but extends to Artiodactyla, specifically within the genus *Nanger*. The presence of both satellite-based and satellite-free centromeres highlights the dynamic nature of centromere evolution in species with highly reshuffled karyotypes. Furthermore, the observation of "epiallelism" or centromere sliding in some loci confirms that centromere position can vary between homologous chromosomes.

In conclusion, this work provides the first comprehensive description of centromere, satellite DNA and telomeric repeats organization in *Nanger dama* and *Nanger soemmerringii*. The centromere architecture in these two species suggests that the absence of satellite DNA at centromeric loci is

more widespread than previously proposed, especially among mammalian species characterized by high genome instability and karyotype plasticity.

ABBREVIATIONS

bp: base pairs

BTA: *Bos Taurus*

CENP: CENtromere Protein

CENP-A: Centromere Protein A

ChIP-seq: Chromatin ImmunoPrecipitation-sequencing

CREST: Calcinosis, Raynaud phenomenon, Esophageal dysmotility, Sclerodactyly, Telangiectasia

ECA: *Equus caballus*

ENC: Evolutionarily New Centromere

FISH: Fluorescent *In Situ* Hybridization

het-ITS: heterochromatic Interstitial Telomeric Sequences

IGV: Integrative Genomics Viewer

ITS: Interstitial Telomeric Sequence

kb: kilobases

Mb: Megabases

Mya: Million years ago

NCBI: National Center for Biotechnology Information

NDA: *Nanger dama*

NSO: *Nanger soemmerringii*

p: short arm of a chromosome

PCR: Polymerase Chain Reaction

RT2T: Ruminant Telomere-to-Telomere

TAREAN: TAndem REpeat Analyzer

INTRODUCTION

1. THE CENTROMERE

1.1 The centromere: a “paradoxal” *locus*

The centromere is a highly specialized nucleoprotein domain of eukaryotic chromosomes that is essential for the accurate segregation of sister chromatids, thereby ensuring the transmission of genetic material during both mitosis and meiosis (Choo, 1997).

From a cytogenetic perspective, the centromere is traditionally identified as the "primary constriction" on metaphase chromosomes. It acts as the physical foundation for the assembly of the kinetochore, a sophisticated multiprotein complex that facilitates the attachment of spindle microtubules (Choo, 2000) (**Figure 1**). This connection is vital for the balanced distribution of the genome into daughter cells.

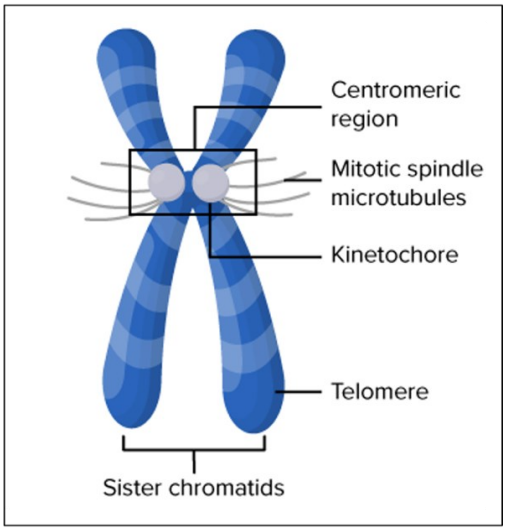

Figure 1. Schematic representation of a metaphase chromosome and the centromere-kinetochore interface. The diagram illustrates the centromeric region, traditionally identified as the primary constriction, where sister chromatids are associated. The kinetochore is depicted as the specialized multiprotein assembly that mediates the anchoring of mitotic spindle microtubules, a fundamental interaction for accurate chromosome alignment and subsequent segregation (Adapted from Choo, 2000).

The functional integrity of the centromere is indispensable for the maintenance of genomic stability and overall cell survival. Indeed, any centromere malfunction or incorrect kinetochore assembly can lead to catastrophic chromosome imbalances, known as aneuploidy. In mammals, such errors are recognized as primary drivers of spontaneous miscarriages, congenital disorders, and infertility (Hassold and Hunt, 2001; Allshire & Karpen, 2008). Furthermore, chromosomal instability resulting

from defective centromeres is frequently observed in cancer, where it contributes to the activation of oncogenes or the loss of tumor suppressor genes (Amor *et al*., 2004; Weaver and Cleveland, 2007; Thompson *et al*., 2010).

A striking observation in centromere biology is that, while its fundamental function is remarkably conserved across diverse evolutionary taxa, the underlying DNA sequences exhibit extreme variability. Centromeric sequences diverge rapidly not only between different species but even between different chromosomes within the same organism. This conceptual tension—conserved function versus divergent sequence—has been termed the "centromere paradox" (Henikoff *et al*., 2001), which is explained by the fact that centromeric identity is not dictated by the underlying DNA sequence but is instead established through epigenetic factors (Henikoff *et al*., 2001).

The primary epigenetic determinant of centromeric chromatin is the presence of a specialized histone H3 variant known as CENP-A (Centromere Protein A, or CID in *Drosophila*). Unlike most genomic regions defined by their underlying DNA sequence, centromere identity in virtually all eukaryotes is specified epigenetically by this histone variant, which replaces canonical H3 within centromeric nucleosomes (Allshire & Karpen, 2008). CENP-A acts as the essential molecular marker that distinguishes active centromeres from the rest of the genome and serves as the foundation for kinetochore recruitment (Henikoff *et al*., 2001; Smith, 2002; Piras *et al*., 2010; Fukagawa & Earnshaw, 2014). The critical nature of CENP-A is further demonstrated by the consequences of its dysregulation; for instance, its overexpression leads to ectopic incorporation into non-centromeric regions, resulting in the formation of misplaced, functional kinetochores that cause severe genomic instability (Heun *et al*., 2006; Moreno-Moreno *et al*., 2006; Shrestha *et al*., 2017).

1.2 The epigenetic determinant: CENP-A

The functional identity of the centromere is epigenetically defined by the presence of CENP-A (or CenH3), a centromere-specific variant of histone H3 that replaces its canonical counterpart within the histone octamers of centromeric chromatin (Allshire & Karpen, 2008).

Historically, CENP-A was identified as a key centromeric antigen through the study of CREST syndrome, a clinical variant of systemic sclerosis characterized by calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia (Earnshaw & Rothfield, 1985; Bobeica *et al.*, 2022). The sera of these patients contain anti-centromere antibodies (ACA) that recognize several proteins, including CENP-B, CENP-C, and most notably CENP-A, which serves as the fundamental marker for the epigenetic propagation of the centromere (Earnshaw & Rothfield, 1985).

In the human genome, CENP-A is encoded by the *CENPA* gene located on chromosome 2 (2p23.3), which produces two isoforms via alternative splicing: a canonical 140-amino acid protein (~16 kDa) and a shorter 114-residue variant (~13 kDa) (UniProt entry: P49450). Structurally, CENP-A consists of a highly divergent N-terminal tail, which varies significantly across species, and a conserved C-terminal tail flanking a central Histone Fold Domain (HFD) that shares 62% homology with histone H3 (Sullivan *et al.*, 1994; UniProt entry: P49450) **(Figure 2).** A pivotal feature within the HFD is the CENP-A Targeting Domain (CATD), which consists of the L1 loop and the second a-helix (2). The CATD is both necessary and sufficient for the precise targeting of the protein to the centromere, ensuring its exclusive localization and the subsequent assembly of the centromeric machinery (Black *et al*., 2007; Goutte-Gattat *et al*., 2013).

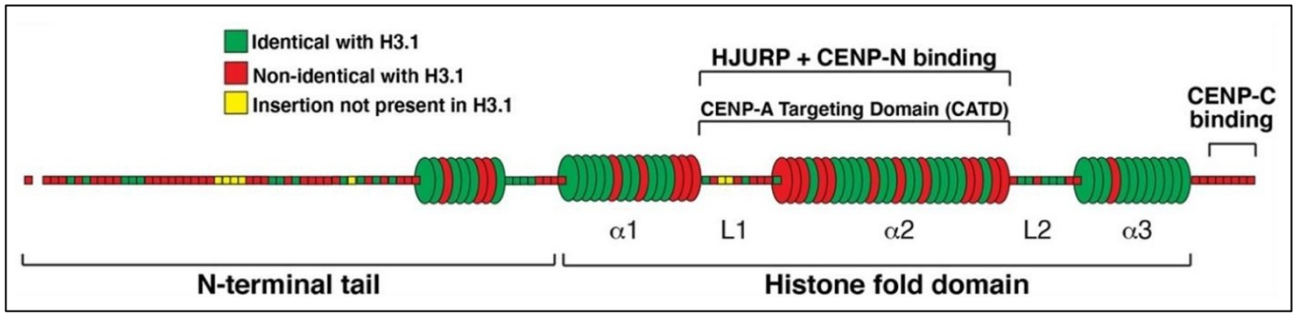

Figure 2. Structural organization and sequence comparison of human CENP-A and histone H3.1. Schematic representation of CENP-A domains, where residues are color-coded based on their relationship with the canonical histone H3.1: green indicates identity, red highlights non-identical residues, and yellow represents insertions unique to CENP-A. The protein features a highly divergent N-terminal tail and a central Histone Fold Domain (HFD) consisting of three a-helices (a1-a3) and two loops (L1, L2). The CENP-A Targeting Domain (CATD), which spans the L1 loop and the a2 helix, is identified as the essential region for HJURP chaperone recognition and CENP-N binding. Additionally, the C-terminal tail is indicated as the primary site for CENP-C interaction (Adapted from Black *et al.*, 2007; Goutte-Gattat *et al.*, 2013).

The incorporation of CENP-A into the nucleosome creates a structurally rigid and compact chromatin environment that is essential for the recruitment of the Constitutive Centromere Associated Network (CCAN) and the subsequent assembly of the kinetochore (Sekulic *et al.*, 2010; Palmer *et al.*, 1987). This deposition process is strictly regulated and occurs during the G1 phase of the cell cycle, mediated by the specialized chaperone HJURP (Holliday Junction Repair Protein), which specifically binds the CATD of newly synthesized CENP-A/H4 complexes (Perpelescu *et al.*, 2015; de Groot *et al.*, 2021; Foltz *et al.*, 2009). Once deposited, CENP-A nucleosomes are interspersed with blocks of canonical H3 nucleosomes, forming a higher-order structure that exposes the CENP-A surface for interaction with inner kinetochore components such as CENP-N and CENP-C (Blower *et al.*, 2002; Zhou *et al.*, 2022). Specifically, CENP-N recognizes the CATD surface bulge, while CENP-C acts as a

"keystone" protein by binding both the CATD and the C-terminal tail of CENP-A (Carroll *et al.*, 2010; Kixmoeller *et al.*, 2020; Zhou *et al.*, 2019). The biological necessity of CENP-A is evidenced by the fact that its depletion leads to catastrophic mitotic failure and embryonic lethality in knockout models, as seen in *CENP-A* null mice (Howman *et al.*, 2000; Takahashi *et al.*, 2000). Conversely, the overexpression of the *CENPA* gene induces its ectopic incorporation into non-centromeric regions, promoting the formation of functional neocentromeres on multicentric chromosomes, which results in severe genomic instability and chromosome segregation defects (Heun *et al.*, 2006; Allshire & Karpen, 2008). This dynamic—where CENP-A is present only at active sites and disappears from inactivated ones—confirms that centromere maintenance is a strictly epigenetic process (Henikoff *et al.*, 2001; Warburton *et al.*, 1997; Earnshaw & Migeon, 1985).

1.3 Centromere-associated proteins and kinetochore assembly

The functional architecture of the centromere is established through a complex assembly of specialized proteins that, despite the extreme variability of centromeric DNA, exhibit high evolutionary conservation across diverse taxa (Henikoff *et al*., 2001). These centromere-associated proteins (CENPs) are categorized into two primary groups: constitutive proteins, which remain associated with the centromere throughout the entire cell cycle, and transient proteins, which localize to the locus only during specific phases (Przewloka & Glover, 2009; Henikoff *et al*., 2001). The systematic identification of these factors was historically initiated by the study of autoimmune sera from patients with CREST syndrome, whose anti-centromere antibodies (ACAs) allowed for the discovery of the first three centromeric antigens: CENP-A (17 kDa), CENP-B (80 kDa), and CENP-C (140 kDa) (Earnshaw & Rothfield, 1985; Earnshaw *et al*., 1985). Collectively, these proteins facilitate the formation of the kinetochore, a multiprotein structure divided into two distinct functional domains—the inner and outer kinetochore—which together mediate the physical connection between chromosomes and the spindle apparatus (Przewloka & Glover, 2009; Dong & Li, 2022) **(Figure 3).**

The inner kinetochore is primarily composed of the Constitutive Centromere Associated Network (CCAN), a group of 16 subunits organized into five distinct modules: CENP-L-N, CENP-H-I-K-M, CENP-O-P-Q-R-U, CENP-T-W-S-X, and the keystone protein CENP-C (Dong & Li, 2022). The assembly of this network is a highly coordinated, hierarchical process. During the G1 phase of the human cell cycle, the histone chaperone HJURP mediates the deposition of new CENP-A molecules by binding to its Centromere-A Targeting Domain (CATD) (Westermann & Schleiffer, 2013). This same CATD domain subsequently facilitates the recruitment of CENP-N, while the C-terminal tail of CENP-A serves as the docking site for CENP-C (Westermann & Schleiffer, 2013; Zhou *et al*.,

2019). While most CENPs lack sequence-specific DNA binding, CENP-B represents a notable exception as the only centromeric protein with high DNA specificity, recognizing a 17 bp motif known as the CENP-B box (Masumoto *et al*., 1989). Although CENP-B is non-essential for centromere function—as evidenced by clinical neocentromeres and Y chromosomes—it is proposed to enhance segregation fidelity by promoting the recruitment of other CCAN components like CENP-C (Fachinetti *et al*., 2015; Chardon *et al*., 2022). Interestingly, recent research in equids has revealed numerous centromeres naturally devoid of CENP-B, suggesting that the dissociation of this protein from the centromere may be a driver of karyotype plasticity (Cappelletti *et al*., 2025).

Bridging the inner kinetochore to the mitotic spindle is the outer kinetochore, which is defined by the KMN network (Knl1-Mis12-Ndc80). This network consists of 10 protein subunits further subdivided into three subcomplexes: the Ndc80 complex, which acts as the direct receptor for spindle microtubules; the Mis12 complex, which serves as a scaffold linking the outer kinetochore to inner components like CENP-C and CENP-T; and the Knl1 complex, which is critical for monitoring correct spindle assembly (Carroll *et al*., 2010; Dong & Li, 2022). The assembly of the CCAN is characterized by a significant degree of dynamism, with protein abundance and subunit interactions shifting throughout the cell cycle to accommodate the requirements of chromosome alignment (Dong & Li, 2022). Evolutionarily, the KMN network is remarkably conserved across eukaryotes, whereas the CCAN components show greater variability between species, reflecting the diverse epigenetic landscapes on which these conserved mitotic machines must operate (Hamilton & Davis, 2020). Through the coordinated folding of higher-order chromatin, CENP-A nucleosomes are positioned to expose these protein networks, ultimately ensuring the maintenance of genomic integrity during every round of division (Blower *et al*., 2002; Palmer *et al*., 1987).

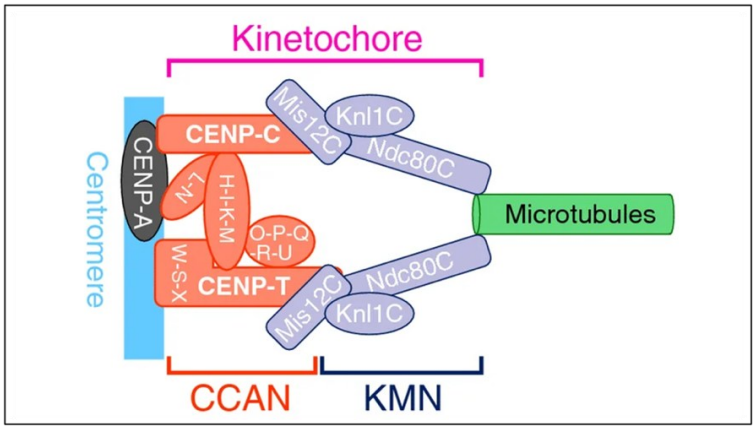

Figure 3. Structural organization of the human kinetochore. The diagram illustrates the protein hierarchy that mediates the connection between centromeric DNA and the microtubules of the mitotic spindle. The inner kinetochore is composed of the CCAN (Constitutive Centromere Associated Network), which specifically recognizes CENP-A-containing nucleosomes. This network serves as a platform for the assembly of the outer kinetochore, or KMN network (Knl1-Mis12-Ndc80), which directly interacts with microtubules. Key proteins such as CENP-C and CENP-T bridge the two domains, ensuring the structural integrity required for accurate sister chromatid segregation (Adapted from Dong & Li, 2022; Dong *et al.*, 2007).

1.4 Centrochromatin: epigenetic identity and transcriptional activity

The functional centromere is defined by a unique chromatin architecture, frequently referred to as "centrochromatin,"which represents a hybrid state distinct from both canonical euchromatin and heterochromatin (Blower *et al*., 2002; Sullivan & Karpen, 2004; Lam *et al*., 2006). At the structural level, nucleosomes containing CENP-A possess a characteristic configuration: the (CENP-A-H4) tetramer is significantly more rigid and compact than the canonical (H3-H4) tetramer, a property directly conferred by the CATD domain (Black et al, 2007; Sekulic *et al*., 2010). Paradoxically, *in vitro* studies indicate that these CENP-A-containing nucleosomes may disassemble more readily than those containing histone H3; this increased turnover is thought to be a specialized feature that facilitates the removal of CENP-A from non-centromeric, ectopic loci, thereby preventing the formation of deleterious neocentromeres (Allshire & Karpen, 2008).

High-resolution immunofluorescence analysis of interphase chromatin fibers has revealed that centrochromatin is characterized by a specific "epigenetic barcode" (**Figure 4**). The centromeric core is highly enriched in H3K4me2 (dimethylation of lysine 4 on histone H3) and H3K36me2, modifications typically associated with an open and permissive chromatin state (Sullivan & Karpen, 2004; Bergmann *et al*., 2011). However, it remains distinct from active euchromatin as it lacks definitive marks such as H3K4me3 or H3K9ac (Naughton & Gilbert, 2020; Lam *et al*., 2006). Furthermore, the integrity of this domain requires H2B monoubiquitination, the depletion of which

leads to inappropriate heterochromatization of the core and subsequent chromosome segregation errors (Sadeghi *et al*., 2014).

This permissive core is flanked by pericentromeric domains heavily enriched with repressive heterochromatic marks, including H3K9me2, H3K9me3, and H3K27me3 (Martins *et al*., 2016; Sullivan & Karpen, 2004). These marks serve to recruit Heterochromatin Protein 1 (HP1), a conserved repressor that binds specifically to methylated H3K9 residues via its N-terminal chromodomain (CD), while utilizing its C-terminal chromo-shadow domain (CSD) for dimerization and various protein-protein interactions (Eissenberg *et al*., 1990; Zeng *et al*., 2010). The precise balance between these activating and silent modifications is critical for centromere identity. As demonstrated in studies using Human Artificial Chromosomes (HACs), shifting this equilibrium—either through the depletion of H3K4me2 and RNA Polymerase II or the over-amplification of heterochromatic silencing—results in the gradual loss of CENP-A deposition and the total collapse of centromeric function (Nakano *et al*., 2008).

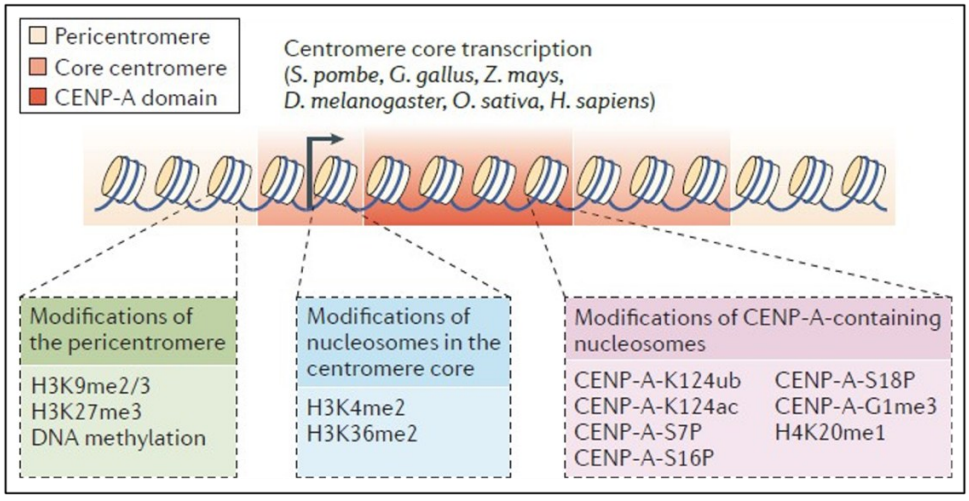

Figure 4. The epigenetic barcode of centromeric and pericentromeric domains. Schematic representation of the histone and DNA modifications defining "centrochromatin." The centromere core and CENP-A domain are characterized by permissive marks (H3K4me2, H3K36me2) and specific CENP-A post-translational modifications (e.g., K124ub, S7P), which support core transcription across multiple species. In contrast, the pericentromere is enriched with repressive marks such as H3K9me2/3 and DNA methylation, which maintain the surrounding heterochromatic state (Adapted from Sullivan & Karpen, 2004; Bergmann *et al.*, 2011).

For decades, the primary constriction of chromosomes was viewed as transcriptionally inert. Early studies in yeast and with Human Artificial Chromosomes (HACs) suggested that high levels of transcription were incompatible with centromere stability (Hill *et al*., 1987; Molina *et al*., 2016). However, recent evidence across various species—from yeast to humans—has overturned this view, demonstrating that both the centromeric core and pericentromeric regions are transcribed at low levels

by RNA Polymerase II (Ohkuni & Kitagawa, 2012; Ling & Yuen, 2019; Hedouin *et al.*, 2022). Investigating this process remains technically challenging due to the highly repetitive nature of centromeric DNA, which complicates the distinction between functional centromeric transcripts and those originating from pericentromeric heterochromatin (Zhu *et al.*, 2023). Nevertheless, centromeric RNAs (cenRNAs) are now recognized as integral components of the centromere, varying in length and abundance and serving essential roles in chromatin remodeling and protein recruitment.

The functional significance of cenRNAs is closely tied to the cell cycle: transcription during interphase is implicated in the chromatin remodeling necessary for the *de novo* deposition of newly synthesized CENP-A in early G1, while mitotic transcription ensures accurate chromosome segregation (Bobkov *et al.*, 2018; Chan *et al.*, 2012). Specifically, both CENP-A and its chaperone HJURP interact directly with cenRNAs to facilitate the proper incorporation of the histone variant into centromeric chromatin (Quenet & Dalal, 2014; Molina *et al.*, 2016). Furthermore, these transcripts play a crucial structural role in stabilizing the kinetochore complex. CENP-C, a fundamental inner kinetochore protein, possesses RNA-binding activity that is necessary for centromere maintenance and stable association with the mitotic spindle (Corless *et al.*, 2020). This link was early evidenced by electron microscopy, which revealed that RNase treatment causes the loss of the electron-dense layer between the centromeric locus and the inner kinetochore (Rieder, 1979). In humans, *Drosophila*, and *Xenopus*, the depletion of specific cenRNAs results in a marked reduction in CENP-C recruitment, further highlighting the universal importance of RNA in defining a functional centromere (Corless *et al.*, 2020).

Despite these findings, the universal requirement for transcription remains a subject of intense debate. The functional plasticity of the centromere is highlighted by the study of neocentromeres, which can emerge and remain stable within both actively transcribed and transcriptionally silent regions of the genome, regardless of the presence of canonical satellite DNA (Saffery *et al.*, 2003; Cardone *et al.*, 2006; Marshall *et al.*, 2008; Piras *et al.*, 2023). Ultimately, while excessive transcriptional activity may disrupt centromere stability, a finely regulated level of RNA synthesis remains a hallmark of functional “centrochromatin”. This transcriptional and epigenetic equilibrium acts as a vital link, bridging the gap between DNA sequence and the assembly of the mitotic machinery (Bobkov *et al.*, 2018; Rošić & Erhardt, 2016).

1.5 Chromosome classification and centromere diversity

Eukaryotic organisms exhibit a wide range of centromeric organizations, which can be broadly categorized into two primary types based on the spatial distribution of their centromeric activity: monocentric and holocentric centromeres (Clarke, 1998; Choo, 2000; Nagaki *et al*., 2005) (**Figure 5)**.

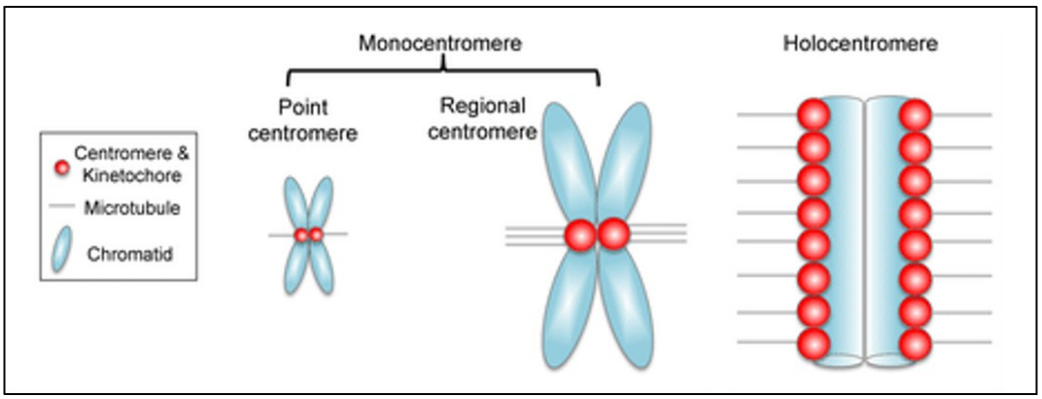

Figure 5. Diversity in centromere organization among eukaryotes. The diagram compares point centromeres (e.g., in *S. cerevisiae*), regional centromeres (found in most plants and animals), and holocentric centromeres (e.g., in *C. elegans*). While point and regional centromeres are localized to a specific chromosomal site, holocentric chromosomes exhibit kinetochore activity and microtubule binding along their entire length.

In the majority of eukaryotes, centromeric function is restricted to a single, discrete locus on the chromosome, a configuration known as a monocentric centromere. These structures are further distinguished into two main classes: point centromeres and regional centromeres.

The point centromere, exemplified by the budding yeast *Saccharomyces cerevisiae*, represents the simplest known centromeric architecture. Spanning approximately 120-125 base pairs, its identity is strictly determined by its DNA sequence, which is both necessary and sufficient to recruit a kinetochore that binds to a single microtubule (Cleveland *et al*., 2003; Malik & Henikoff, 2009). Notably, *S. cerevisiae* remains the only documented exception to the "centromere paradox," as its centromeric function is sequence-dependent rather than epigenetically defined.

In contrast, regional centromeres are found in the vast majority of eukaryotes, including humans, mice, and *Drosophila*. These centromeres occupy significantly larger genomic domains, ranging from tens of kilobases to several megabases, and are typically composed of complex arrays of repetitive satellite DNA and retro-transposable elements (Kalitsis and Choo, 2012; McKinley & Cheeseman, 2016). According to centromere position, monocentric chromosomes can be classified in telocentric, acrocentric, sub-metacentric and metacentric chromosomes (Levan *et al*., 1964), (**Figure 6**).

A fundamentally different organization is observed in holocentric centromeres, where centromeric activity and microtubule attachment are not localized to a single point but are instead distributed along the entire length of the chromosome. This configuration is characteristic of certain plants, insects, and nematodes, such as *Caenorhabditis elegans* (Melters *et al*., 2013; McKinley & Cheeseman, 2016). Because the kinetochore forms along the whole chromatid, these chromosomes lack a visible primary constriction during mitosis. Interestingly, the evolutionary plasticity of this system is highlighted by some holocentric species, such as those in the order Lepidoptera, which have been found to lack the histone variant CENP-A entirely. This suggests that, in specific lineages, alternative chromatin features can occasionally drive faithful chromosome segregation independently of the canonical CENP-A pathway (Senaratne *et al*., 2021, 2022).

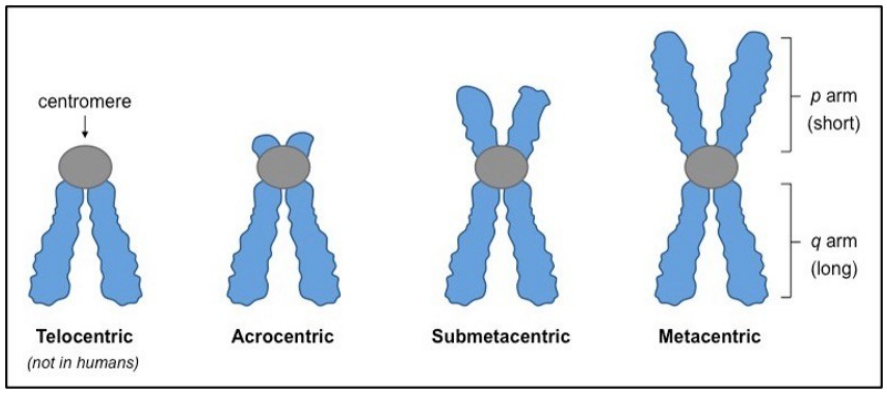

Figure 6. Morphological classification of monocentric chromosomes. Chromosomes are categorized based on the position of the centromere (primary constriction) as metacentric, sub-metacentric, acrocentric, or telocentric, determining the relative length of the p and q arms.

Despite the prevalence of satellite DNA, it has become increasingly clear that these sequences are neither strictly necessary nor sufficient for centromere establishment (Choo, 1997, 2000). This is exemplified by human isodicentric chromosomes—where only one of two identical satellite-rich regions remains active (Sullivan and Willard, 1998) —and the emergence of "neocentromeres" at ectopic, non-satellite loci (Marshall *et al*., 2008; Kalitsis and Choo, 2012). Satellite-less centromeres have also been identified in several species, including the fungus *Candida albicans*, the orangutan, and the chicken (McKinley & Cheeseman, 2016; Legrand *et al*., 2019).
A particularly significant breakthrough in this field was the discovery by our laboratory of a functional, satellite-free centromere on horse chromosome 11, which is fixed in the population (Wade *et al*., 2009). Subsequent research identified numerous other satellite-less centromeres across different species of the genus *Equus* (Piras *et al*., 2010; Nergadze *et al*., 2018; Cappelletti *et al*., 2022; Piras *et al*., 2023). Due to these unique features, equids have emerged as an invaluable model

system for studying the evolution, formation, and epigenetic maintenance of centromeres in the absence of repetitive DNA (Giulotto *et al*., 2017; Piras *et al*., 2022). A more comprehensive characterization of neocentromeres and the specific role of the genus *Equus* as a specialized model system will be discussed in Chapters 2 and 3, respectively.

1.6 Centromeric DNA and its functional organization

Centromeric sequences exhibit remarkable diversity across the eukaryotic domain, reflecting a complex evolutionary history. In the specific case of point centromeres, such as those in *Saccharomyces cerevisiae*, centromeric function is strictly dependent on the underlying DNA sequence. These centromeres are organized into three distinct functional domains: CDE I (8 bp), CDE II (26 bp), and CDE III (78–86 bp). While CDE I and CDE III are recognized by specialized proteins—such as Cbf1 and the CBF3 complex, respectively—CDE III and specific portions of the AT-rich CDE II domain are considered indispensable for recruiting the single microtubule that defines yeast segregation (Lechner & Ortiz, 1996; Malik & Henikoff, 2009).
In contrast, the regional centromeres found in the majority of eukaryotes typically consist of vast blocks of satellite DNA and retrotransposons, with sequences that vary significantly between organisms (Plohl *et al*., 2014). Despite this variability, regional centromeres share a conserved modular architecture. They generally feature a central "core" domain composed of ordered, homogeneous repeats where the histone H3 variant CENP-A is deposited. This core is surrounded by an extensive "pericentromeric" region characterized by more degenerate repeats interspersed with mobile elements. While the pericentromere maintains a repressive heterochromatic environment, the core domain mediates microtubule attachment and ensures proper chromosomal segregation (McKinley & Cheeseman, 2016).
Human centromeres serve as a primary model for this regional organization. Their sequence is dominated by the alpha-satellite family, which comprises approximately 5% of the human genome and consists of AT-rich monomers of 171 bp arranged in tandem head-to-tail arrays (Maio, 1971; Willard, 1991). These monomers are further organized into Higher-Order Repeats (HORs)—chromosome-specific arrays spanning from 250 to 5,000 kilobases—which constitute the functional centromeric core (Shepelev *et al*., 2015; Hartley & O'Neill, 2019) (**Figure 7**). Although satellite DNA was long dismissed as non-functional "junk" DNA, evidence from Human and Yeast Artificial Chromosomes (HACs and YACs) has demonstrated that alpha-satellite arrays are indeed essential for stable inheritance and kinetochore stability (Yunis *et al*., 1971; Harrington *et al*., 1997; Ikeno *et al*., 1998).

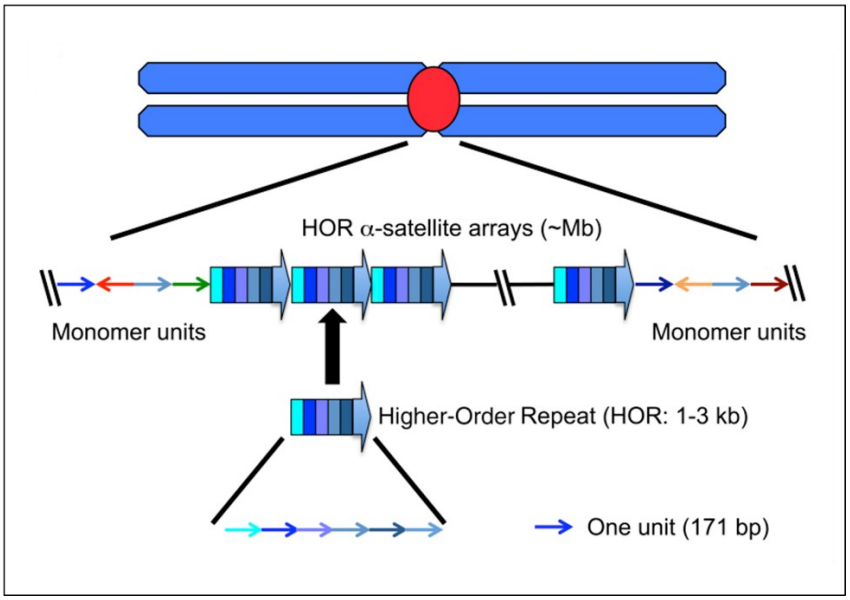

Figure 7. Structural architecture of human centromere. The diagram illustrates the organization of the alpha-satellite family from the chromosomal level to the primary 171 bp monomer unit. Individual monomers are clustered into chromosome-specific Higher-Order Repeats (HORs), which are amplified to form the expanded satellite arrays necessary for kinetochore stability and stable inheritance (Adapted from Hartley & O'Neill, 2019).

The rapid evolution and expansion of these repeats are often explained by the "centromere drive" hypothesis (**Figure 8**). Because female meiosis is asymmetric, centromeres compete for inclusion in the single viable oocyte. "Stronger" centromeres, characterized by larger satellite arrays, can recruit more CENP-A and kinetochore proteins, favoring their transmission (Malik and Bayes, 2006; Kursel & Malik, 2018). To counteract this potentially deleterious expansion, kinetochore proteins like CENP-A and CENP-C co-evolve to suppress this drive, maintaining a balance between repeat expansion and genomic stability (Henikoff *et al*., 2001; Talbert & Henikoff, 2022; Dudka & Lampson, 2022). This rapid divergence is further supported by the "molecular drive" model, involving mechanisms like unequal crossing-over and gene conversion (Dover *et al*., 1982; Elder and Turner, 1995; Garrido-Ramos, 2017), and the "library hypothesis," which suggests that related species differentially amplify specific families from a common ancestral set of satellite sequences (Salser *et al*., 1976; Fry & Salser, 1977). Interestingly, the fact that centromeric monomers often correspond to multiples of the nucleosomal wrap length—171 bp in humans or 178 bp in *Arabidopsis*—suggests that structural constraints like nucleosome positioning also influence their evolution (Henikoff *et al*., 2001; Cleveland *et al*., 2003; Melters *et al*., 2013).

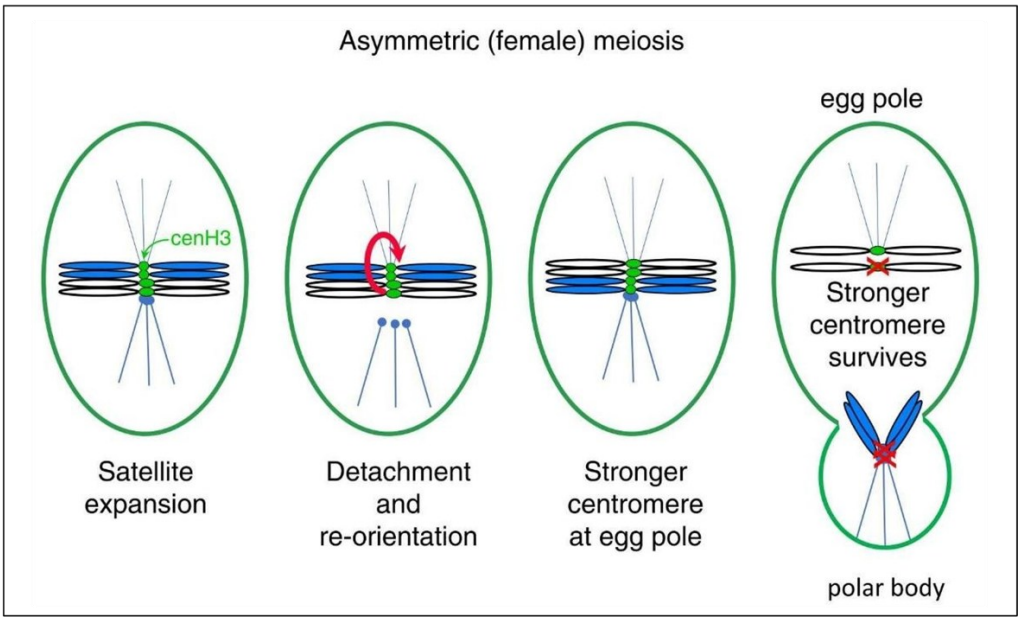

Figure 8. The centromere drive hypothesis in asymmetric female meiosis. Schematic representation of the competitive selection of centromeres during oogenesis. (left to right) Satellite DNA expansion leads to increased recruitment of cenH3 (CENP-A), resulting in a "stronger" centromere. During meiotic division, these stronger centromeres can trigger detachment and re-orientation of the chromosome to preferentially face the egg pole. Consequently, the stronger centromere is retained within the functional oocyte, while the weaker centromere is segregated into the polar body and subsequently discarded (Adapted from Malik and Bayes, 2006; Kursel & Malik, 2018).

A notable exception to the rapid divergence is the 17-bp "CENP-B box," the only conserved sequence capable of specifically binding a centromeric protein (CENP-B), which may stabilize the centromere and enhance CENP-A recruitment (Masumoto *et al.*, 1989; Fachinetti *et al.*, 2015; Gamba & Fachinetti, 2020). Ultimately, the discovery of numerous functional, satellite-free centromeres—particularly the 84 loci identified across eight equid species by our laboratory—conclusively demonstrates that satellite DNA is not a universal requirement, positioning the genus *Equus* as a unique model for understanding the birth and epigenetic maintenance of mammalian centromeres (Wade *et al.*, 2009; Piras *et al.*, 2010; Nergadze *et al.*, 2018; Piras *et al.*, 2022).

2. NEOCENTROMERES

2.1 Definition and the process of “centromerization”

Cumulative evidence has demonstrated that although most centromeres are associated with highly repetitive satellite DNA, centromeric function is not strictly determined by the underlying sequence. The most compelling proof for this epigenetic nature is the existence of natural centromeres that completely lack satellite DNA, yet are fully capable of forming primary constrictions and recruiting the multiprotein kinetochore complex (Amor & Choo, 2002).

These loci arise through a phenomenon termed "centromerization" (Choo, 2000). This term encompasses two distinct events: the maintenance of centromeric identity at a specific chromosomal position across cell cycles, and the *de novo* formation of a functional centromere at an ectopic location previously devoid of such function—a structure known as a neocentromere (Choo, 2000). Neocentromeres are broadly classified into two categories: clinical neocentromeres, which appear sporadically in the human population and are not typically inherited, and evolutionarily new centromeres (ENCs), which are stable, functional, and fixed across a species, thereby playing a fundamental role in karyotype evolution and speciation (Rocchi *et al.*, 2012; Piras *et al.*, 2022).

2.2 Human clinical neocentromeres

Clinical neocentromeres have been identified in individuals with chromosomal aberrations linked to congenital diseases or neoplastic mutations. The first instance was documented on the long arm of chromosome 10 (10q25) in a patient with mild developmental delay; despite the absence of a-satellite DNA and CENP-B binding, this "marker chromosome" (mar del (10)) successfully recruited essential centromeric proteins (Voullaire *et al.*, 1993).

The emergence of these neocentromeres often serves as a rescue mechanism for acentric fragments generated by chromosomal rearrangements or breakages during meiosis or mitosis (Depinet *et al.*, 1997; Voullaire *et al.*, 1999) (**Figure 9a**). By acquiring neocentromeric activity, these fragments avoid being lost during cell division. While no specific DNA sequence is required for this process, neocentromerization does not occur at random. Instead, it favors specific "hotspots"—often located in gene-poor, euchromatic distal regions characterized by high LINE (Long Interspersed Nuclear Elements) and AT-rich content (Marshall *et al.*, 2008; Kalitsis & Choo, 2012). Frequent hotspots have been identified at 3q, 8p, 13q, 15q, and Yq (Marshall *et al.*, 2008).

Notably, neocentromeres can also form on structurally intact chromosomes even when the native centromere remains functional, resulting in pseudodicentric chromosomes that are generally non-pathogenic and occasionally discovered in healthy individuals (Marshall *et al.*, 2008; Hasson *et al.*, 2011) (**Figure 9b**).

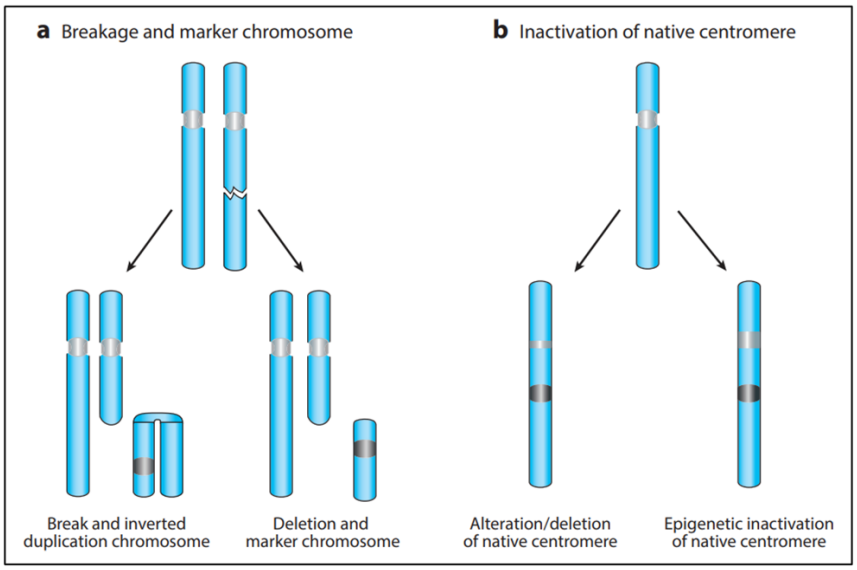

Figure 9. Mechanisms of human neocentromere formation. (a) Neocentromeres can emerge following chromosomal breakage events, such as terminal or interstitial deletions during mitosis or meiosis. These occurrences lead to the generation of supernumerary marker chromosomes (SMCs), which may consist of the excised chromosomal fragment or arise through complex inverted duplication events. **(b)** A second class of neocentromeres forms at ectopic, non-canonical sites on structurally intact chromosomes. This process may serve to rescue a compromised or partially deleted endogenous centromere, involving a phase of competition between the two loci until the epigenetic inactivation of one site is achieved (Adapted from DeBose-Scarlett & Sullivan, 2021).

2.3 Evolutionarily New Centromeres (ENCs)

Unlike clinical cases, ENCs are fixed within a species' genome and arise through centromere repositioning—an epigenetic shift of the centromeric locus along a chromosome without accompanying sequence rearrangements (Montefalcone *et al.*, 1999). This phenomenon was first characterized during the study of primate chromosome IX evolution (Montefalcone *et al.*, 1999). Interestingly, the genomic regions where human clinical neocentromeres are seeded often correspond to the orthologous positions of ENCs in other primates, suggesting that certain domains possess an innate predisposition for centromere formation due to specific chromatin environments (Cardone *et al.*, 2006; Capozzi *et al.*, 2008; Rocchi *et al.*, 2012).

While early primate ENCs were associated with satellite DNA, a landmark discovery identified the first satellite-free evolutionary neocentromere on chromosome 11 (ECA11) of the domestic horse (*Equus caballus*) (Wade *et al.*, 2009). Subsequent research has identified numerous satellite-free centromeres across the genus *Equus*, including donkeys, zebras, and Przewalski's horses (Piras *et al.*, 2010; Nergadze *et al.*, 2018; Cappelletti *et al.*, 2022; Piras *et al.*, 2023). Outside of equids, satellite-free ENCs have been found in the orangutan—where they exist in a polymorphic state alongside ancestral satellite centromeres (Tolomeo *et al.*, 2017) —in the eastern hoolock gibbon (Hartley *et al.*, 2025), and in chickens (Shang *et al.*, 2010). However, the high frequency of such centromeres in equids makes them a unique model system for centromere biology (Giulotto *et al.*, 2017; Cappelletti *et al.*, 2019; Peng *et al.*, 2021; Piras *et al.*, 2022).

3. THE GENUS *EQUUS*: a model system for ENCs study

3.1 The order Perissodactyla

The mammalian superorder Ungulata, a term derived from the Latin *ungulatum* meaning "hoofed," encompasses several orders, most notably the Perissodactyla (odd-toed ungulates) and the Artiodactyla (even-toed ungulates) (**Figure 10**). Within the Perissodactyla, two primary suborders are recognized: Ceratomorpha, which includes the families Tapiridae and Rhinocerotidae, and Hippomorpha, represented today solely by the family Equidae. Paleontological and molecular evidence suggests that the divergence between these lineages occurred approximately 56 million years ago (Mya) during the early Eocene (Trifonov *et al.*, 2008). While the ancestral Perissodactyla once comprised thirteen families, evolutionary pressures have reduced this diversity to the three extant families observed today. The Tapiridae family consists of four species (*Tapirus indicus*, *T. terrestris*, *T. pinchaque*, and *T. bairdii*), while the Rhinocerotidae includes five species (*Diceros bicornis*, *Ceratotherium simum*, *Rhinoceros unicornis*, *Dicerorhinus sumatrensis*, and *Rhinoceros sondaicus*). In contrast, the Equidae family is characterized by a single genus, *Equus*, which encompasses nine species and subspecies, including horses (*Equus caballus* and *E. przewalskii*), asses (*E. asinus*, *E. africanus somaliensis*, and the Asiatic *E. hemionus*), and various zebra species such as *E. zebra*, *E. burchelli*, and *E. grevyi* (Piras *et al.*, 2010; Cappelletti *et al*., 2022).

The evolutionary history of these groups is marked by a profound disparity in chromosomal stability. The diploid chromosome number (2n) exhibits extreme variability, ranging from 52 to 80 in tapirs and 82 to 84 in rhinoceroses, while equids show values between 32 and 66. Despite these numbers, the karyotypes of Ceratomorpha have remained remarkably stable throughout their evolution. With the exception of *Tapirus indicus*, these species retain a chromosomal arrangement that closely resembles the putative ancestral perissodactyl karyotype, characterized by a high frequency of acrocentric chromosomes (Trifonov *et al.*, 2008, 2012). Conversely, the lineage leading to the genus *Equus* underwent an exceptionally rapid evolution following its divergence from the common perissodactyl ancestor approximately 4–4.5 Mya (Orlando *et al.*, 2013). This rapid radiation, which saw the most recent divergence between asses and zebras occurring roughly 1 Mya, was accompanied by extensive genomic remodeling and chromosomal rearrangements (Trifonov *et al.*, 2008; Jónsson *et al.*, 2014).

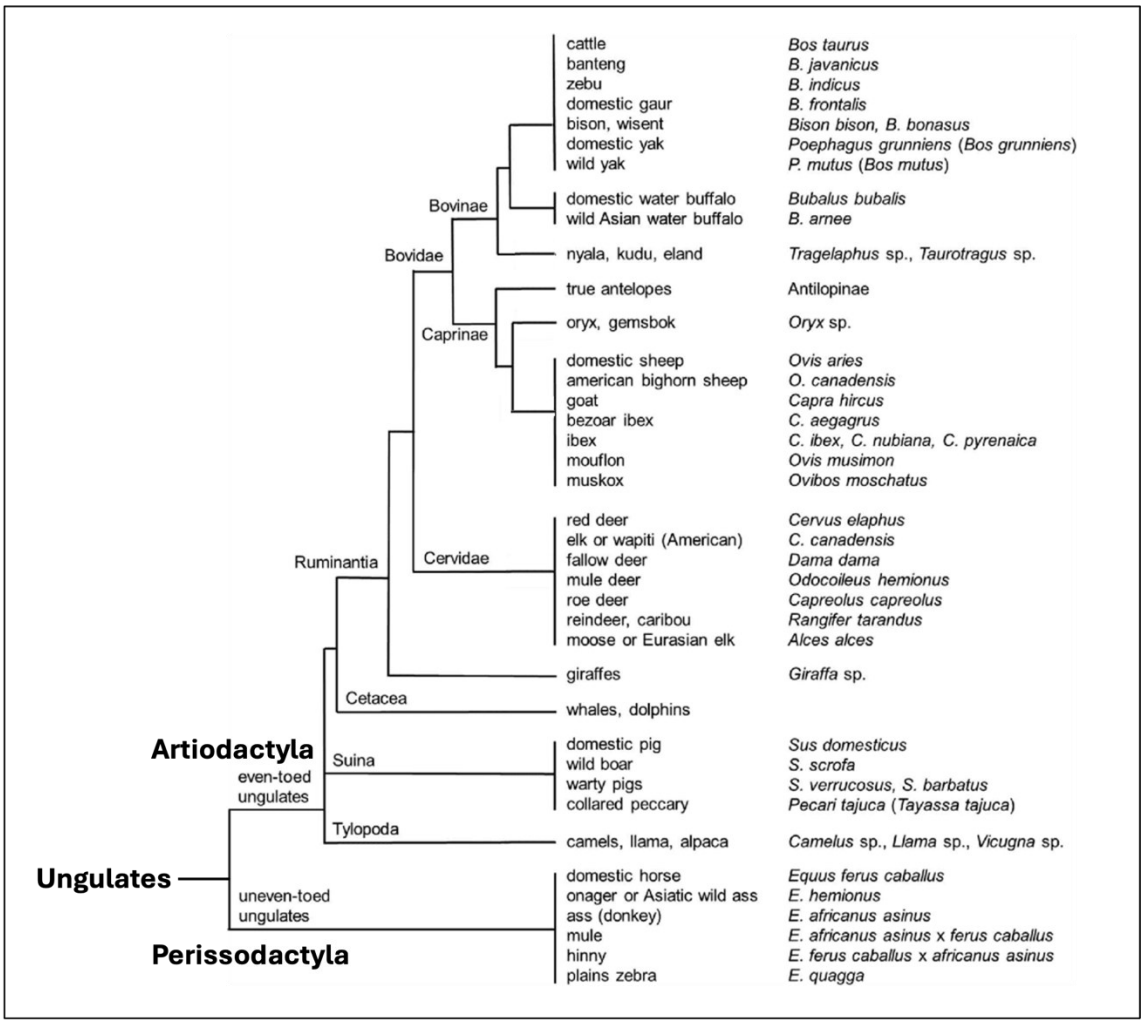

Figure 10. Phylogenetic tree of the superorder Ungulata and related mammalian orders. The cladogram illustrates the evolutionary relationships between the Perissodactyla (odd-toed ungulates) and Artiodactyla (even-toed ungulates and cetaceans) (Adapted from van Raamsdonk *et al*., 2019).

3.2 Neocentromeres in the genus *Equus*

A defining feature of this rapid equid radiation is the frequent occurrence of centromere repositioning and centric fusions, which have collectively facilitated the emergence of numerous evolutionary neocentromeres. Unlike canonical centromeres, these newly formed loci are characterized by a complete absence of satellite DNA arrays (Wade *et al.*, 2009; Piras *et al.*, 2010; Nergadze *et al.*, 2018). The first natural, satellite-free mammalian centromere fixed within a species was discovered by our laboratory on horse chromosome 11 (ECA11) following the sequencing of the *Equus caballus* genome (Wade *et al.*, 2009).

Building upon the discovery of the ECA11 neocentromere, our research group has identified an extraordinary abundance of satellite-free centromeres across other equid species through a combination of cytogenetic approaches, such as FISH with total genomic DNA and major satellite

families, and molecular techniques like ChIP-seq (Piras *et al.*, 2010; Nergadze *et al.*, 2018) (**Figure 10**). In the domestic donkey (*Equus asinus*), 16 out of 31 chromosomes possess satellite-free centromeres, while Burchell's zebra (*E. burchelli*) and Grevy's zebra (*E. grevyi*) harbor 15 out of 22 and 13 out of 23 such centromeres, respectively (Nergadze *et al.*, 2018; Cappelletti *et al.*, 2022). Przewalski's horse (*E. przewalskii*) also retains a satellite-free neocentromere (Piras *et al.*, 2023). Although centromere function is epigenetically determined and theoretically sequence-independent, these equid neocentromeres often reside within specific genomic environments characterized by a high AT content and an enrichment of LINE-1 retrotransposons (Allshire & Karpen, 2008; Nergadze *et al.*, 2018; Cappelletti *et al.*, 2022).

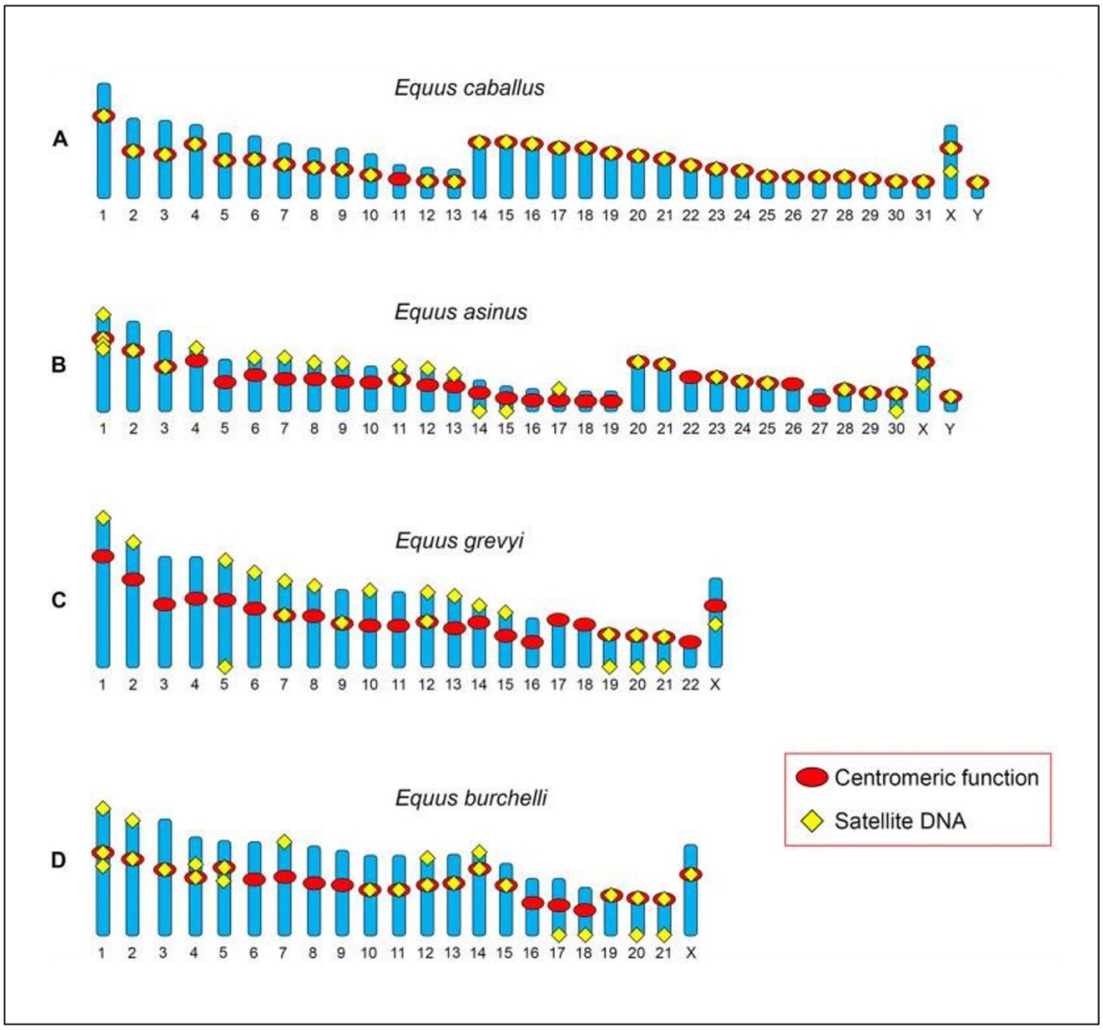

Figure 10. Schematic chromosomal distribution of satellite DNA across four *Equus* species. Idiograms representing the karyotypic arrangements of (**A**) the domestic horse, (**B**) the donkey, (**C**) Grevy's zebra, and (**D**) Burchell's zebra. The diagrams illustrate the heterogeneous centromeric landscape of these species, highlighting the coexistence of canonical satellite-based centromeres and satellite-free functional domains within the same genome (Adapted from Piras *et al.*, 2022).

3.3 Origins and dynamic properties of Equid neocentromeres

The emergence of satellite-free centromeres in the equid lineage is primarily attributed to centromere repositioning (Carbone *et al.*, 2006; Piras *et al.*, 2010) (Route A, **Figure 11**). This process allows for the seeding of a new centromere at an interstitial site, a phenomenon also observed in rare, non-pathogenic human neocentromeres (Rocchi *et al.*, 2012). In our most recent model these satellite-free centromeres are considered an "immature" stage of centromere evolution. While the new centromere operates at a single-copy DNA site, the ancestral satellite DNA may still be visible at the site of the previously active, now-inactivated centromere. Over evolutionary time, these new centromeres may progressively "mature" by acquiring new satellite repeats through mechanisms like DNA amplification.

However, recent research on Burchell's and Grevy's zebras has introduced an additional mechanism: centric fusions (or Robertsonian translocations) (Route B, **Figure 11**). In these species, several satellite-free centromeres were generated following the fusion of ancestral acrocentric chromosomes, where centromeric function was established directly at the fusion site, often utilizing pericentromeric remnants from inactivated ancestral centromeres (Cappelletti *et al.*, 2022).

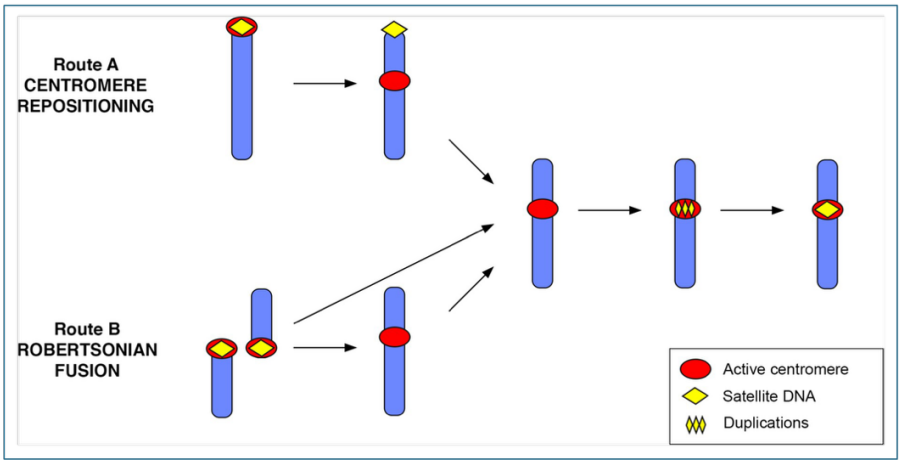

Figure 11. Model for the birth and maturation of centromeres during Equus evolution. Satellite-free centromeres are established through two primary mechanisms: epigenetic repositioning (Route A) or Robertsonian translocations (Route B). Following their initial seeding at these ectopic loci, the new centromeres undergo a progressive maturation process characterized by the gradual recruitment and expansion of satellite DNA arrays (Modified from Piras *et al.*, 2010; Nergadze *et al.*, 2018; Cappelletti *et al.*, 2022).

Beyond their origin, these neocentromeres exhibit a unique degree of positional flexibility. On horse chromosome 11 (ECA11), molecular investigations revealed that the CENP-A binding domain is not anchored to a static sequence but can relocate within a 500 kb genomic window, a process known as centromere sliding (Purgato *et al.*, 2015) (**Figure 12a**). This mobility results in the existence of

positional alleles, or epialleles, within the population (**Figure 12b**). In individuals where homologous chromosomes carry CENP-A binding domains in different locations, molecular analysis identifies two distinct peaks of protein recruitment, reflecting a form of functional allelism without underlying sequence changes (Purgato *et al.*, 2015).

These configurations are inherited as Mendelian traits, yet they remain dynamic. Studies in horse-donkey hybrids have shown that centromere position can slide between 50 and 80 kb across a single generation during parent-to-offspring transmission (Nergadze *et al.*, 2018). The fact that CENP-A domains move within restricted boundaries suggests that centromeric function is physically constrained by local chromatin architecture or specific epigenetic marks (Sullivan & Karpen, 2004; Martins *et al.*, 2016). The presence of these "immature" centromeres, combined with ancestral satellite relics at non-centromeric positions, identifies the genus *Equus* as an unparalleled model for understanding the plasticity and evolutionary maturation of the mammalian centromere (Piras *et al.*, 2010; Giulotto *et al.*, 2017; Piras *et al.*, 2022).

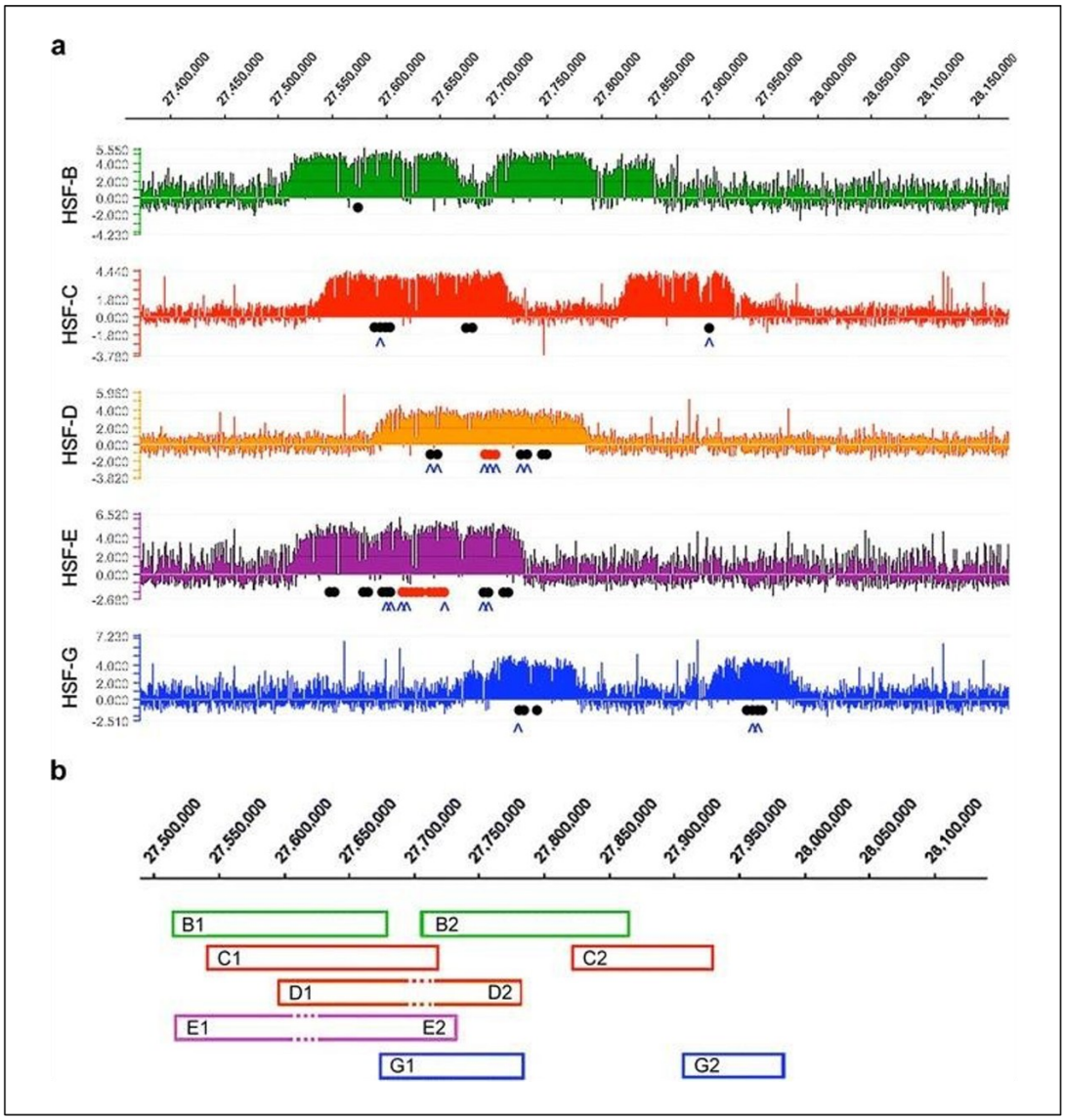

Figure 12. Centromere sliding within the ECA11 centromeric domain. (a) ChIP-on-chip profiles illustrating CENP-A enrichment domains across five domestic horse individuals (HSFD-B, -C, -D, -E, and -G). The enrichment peaks were identified by hybridizing chromatin-immunoprecipitated DNA (anti-CENP-A) and total input DNA to a custom genomic array covering the ECA11 centromeric region. The x-axis indicates genomic coordinates on horse chromosome 11 (ECA11), while the y-axis represents the $\log_2$ ratio of the hybridization signals (ChIP/Input). **(b)** Schematic representation of the positional epialleles identified within the analyzed horse individuals (Adapted from Purgato *et al.*, 2015).

4. THE GENUS *NANGER:* a new model system

4.1 The order Artiodactyla and the gazelles of the genus *Nanger*

In addition to the Perissodactyls, the superorder Ungulata includes the highly diverse order Artiodactyla, or even-toed ungulates (**Figure 10**). A significant majority of artiodactyl species—approximately two-thirds—belong to the suborder Ruminantia, which is primarily divided into two major families: Cervidae and Bovidae. Within the Bovidae family, the subfamily Antilopinae stands out as a critical group for evolutionary studies, particularly the genus *Nanger*. This genus comprises several prominent gazelle species, including *Nanger dama*, *Nanger soemmerringii*, and *Nanger granti*.

The internal classification of these species has been a subject of significant scientific revision. Traditionally, *Nanger dama* was categorized into three distinct subspecies—*N. dama ruficollis*, *N. dama dama*, and *N. dama mhorr*—based on phenotypic variations in coat coloration. However, recent comprehensive molecular and genetic analyses have challenged this division, suggesting that *Nanger dama* should be considered a single, unified species without further sub-classification (Senn *et al.*, 2014). In contrast, *Nanger soemmerringii* remains formally divided into three subspecies characterized by distinct geographical distributions and morphological traits: *N. soemmerringii soemmerringii*, *N. soemmerringii berberana*, and *N. soemmerringii butteri* (Steiner *et al.*, 2015).

From a cytogenetic perspective, gazelles exhibit a remarkably variable diploid chromosome number (2n), ranging from 30 to 58. Comparative analyses across various bovid lineages indicate that the modern gazelle karyotypes were primarily shaped by multiple centric fusions (or Robertsonian translocations) involving the 58 acrocentric chromosomes of the putative ancestral bovid karyotype (Vassart *et al.*, 1995). Furthermore, the genomes of these species are characterized by a high degree of individual variability. This chromosomal plasticity is driven by polymorphic translocations within populations, involving both autosomes and sex chromosomes, which contribute to a state of significant genomic and karyotypic instability (Steiner *et al.*, 2015).

Given the rapid evolutionary radiation and the pronounced genomic instability that mirror the patterns observed in equids, the genus *Nanger* represents an ideal candidate for further centromere research. Consequently, our laboratory has extended the investigation of satellite-free neocentromeres to *Nanger dama* and *Nanger soemmerringii*. By exploring these species, we aim to establish a new potential model system to better understand the epigenetic mechanisms of centromere seeding and the evolutionary dynamics of "centrochromatin" in a broader mammalian context.

4.2 *Nanger Dama* cell lines

The Dama gazelle (*Nanger dama*), also known as the addra gazelle or mhorr gazelle, is a critically endangered species of gazelle belonging to the family Bovidae and the subfamily Antilopinae. Historically distributed across the Sahara and Sahel regions, its populations have undergone a dramatic decline, leaving it as a primary focus for both conservation and cytogenetic research. As with other members of the Antilopinae subfamily, the karyotypic evolution of *Nanger dama* is characterized by a significant reduction in chromosome number compared to the putative ancestral bovid karyotype (2n = 58) which consisted entirely of acrocentric chromosomes (Vassart *et al.*, 1995). The chromosomal arrangement of *Nanger dama* is particularly complex due to the presence of multiple centric fusions and specialized sex-chromosome translocations. In the specific cell line analyzed in our laboratory (Sample ID: SAZ23, derived from the San Antonio Zoo), the diploid number was determined to be 2n = 39 in a male individual. This odd diploid number is a hallmark of the species' sex-determination system, which involves an X-autosome translocation. According to literature, the female of the species typically presents 2n = 40, while the male possesses 2n = 39 due to the fusion of an ancestral autosome to the X chromosome, resulting in a trivalent configuration during meiosis (Vassart *et al.*, 1993; Vassart *et al.*, 1995).

As illustrated in **Figure 13** (adapted from Vassart *et al.*, 1995), the *Nanger dama* karyotype is dominated by large (sub)metacentric chromosomes resulting from Robertsonian translocations (centric fusions). The specific SAZ23 line is characterized by a complex sex chromosome system (X/A; Y/A), where autosomal segments have been incorporated into the sex chromosomes, further contributing to the genomic instability and the unique centrochromatin landscape of this species. The lack of traditional satellite DNA at some of these recent fusion sites makes *Nanger dama* a promising candidate for investigating the seeding of neocentromeres within the Artiodactyla order.

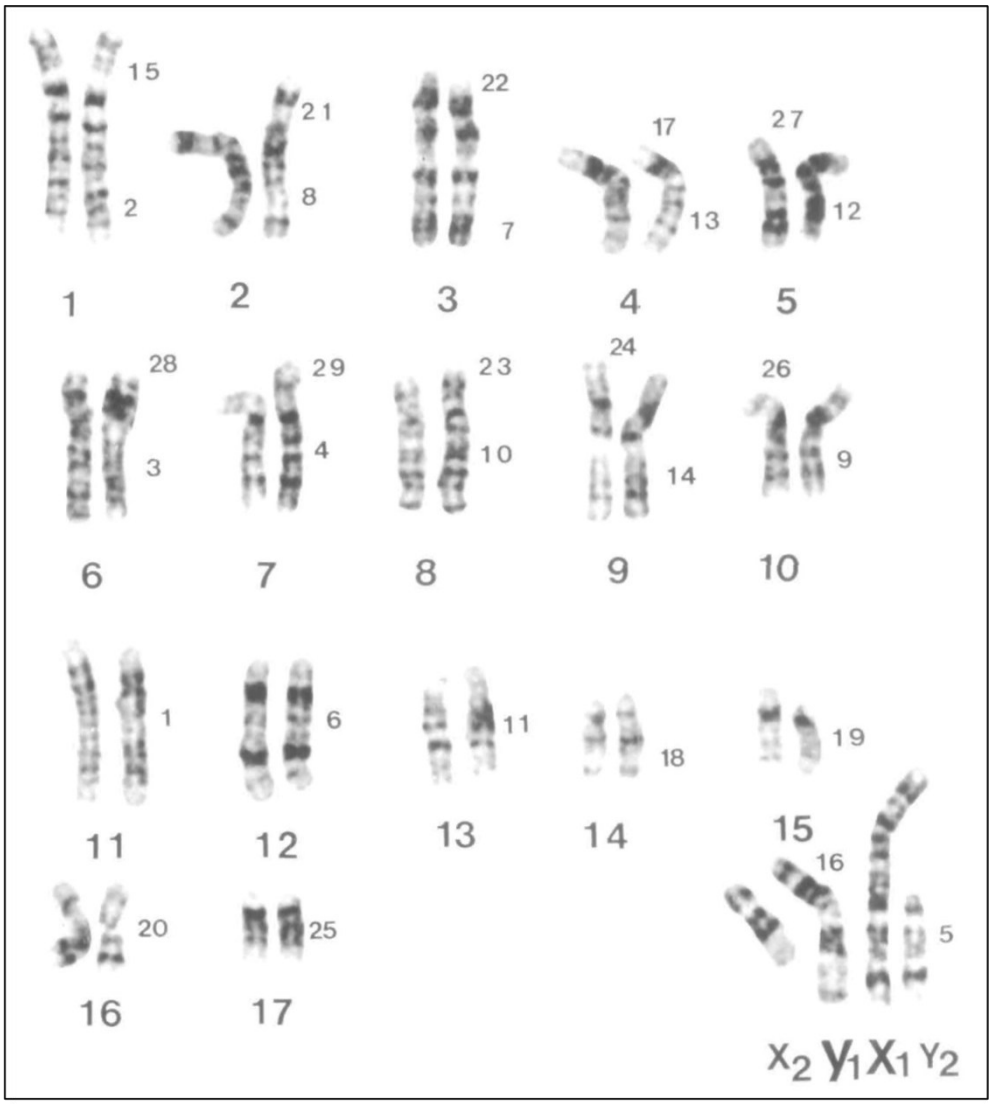

Figure 13. Representative karyotype of *Nanger dama* (2n = 39, male). This arrangement serves as the reference for the SAZ23 cell line (San Antonio Zoo). The karyotype is dominated by large (sub)metacentric chromosomes resulting from centric fusions of ancestral acrocentrics. The presence of X/A and Y/A translocations highlights the genomic plasticity of the Dama gazelle, providing a unique framework for the study of neocentromere formation (Adapted from Vassart *et al.*, 1995).

4.3 *Nanger soemmerringii* cell lines

Soemmerring's gazelle (*Nanger soemmerringii*) is a species native to the Horn of Africa, categorized into three subspecies—*N. s. soemmerringii*, *N. s. berberana*, and *N. s. butteri*—distinguished by their geographical distribution and morphological markers (Steiner *et al.*, 2015). From a cytogenetic standpoint, this species is renowned for its high degree of chromosomal polymorphism, which frequently results in different diploid numbers and varying karyotypic configurations even within the same population (Vassart *et al.*, 1995; Steiner *et al.*, 2015).

In our study, we characterized a specific male cell line (Sample ID: NSO3, originating from the St. Louis Zoo). The reconstructed karyotype for this individual revealed a diploid number of 2n = 37 (**Figure 14**). This specific chromosomal count is lower than the species' typical range and is the result of several fixed and polymorphic centric fusions. Specifically, molecular and cytogenetic analysis identified two critical translocations: a homozygous fusion between chromosomes 11 and 16 (11/16 +/+) and a heterozygous fusion between chromosomes 10 and 18 (10/18 +/-).

The presence of the 10/18 (+/-) polymorphism is particularly significant, as it demonstrates the ongoing karyotypic plasticity within the species. In this heterozygous state, one set of chromosomes remains as independent acrocentrics, while the homologous set has undergone a Robertsonian fusion to form a single (sub)metacentric element. This dynamic state of "centromeric transition" provides a unique opportunity to study the molecular environment of newly formed centromeres. Like *Nanger dama*, the Soemmerring's gazelle karyotype represents a specialized "laboratory of evolution," where the frequent reorganization of the genome facilitates the emergence of neocentromeric domains and the potential dissociation of centromeric function from canonical satellite DNA (Vassart *et al.*, 1995; Steiner *et al.*, 2015).

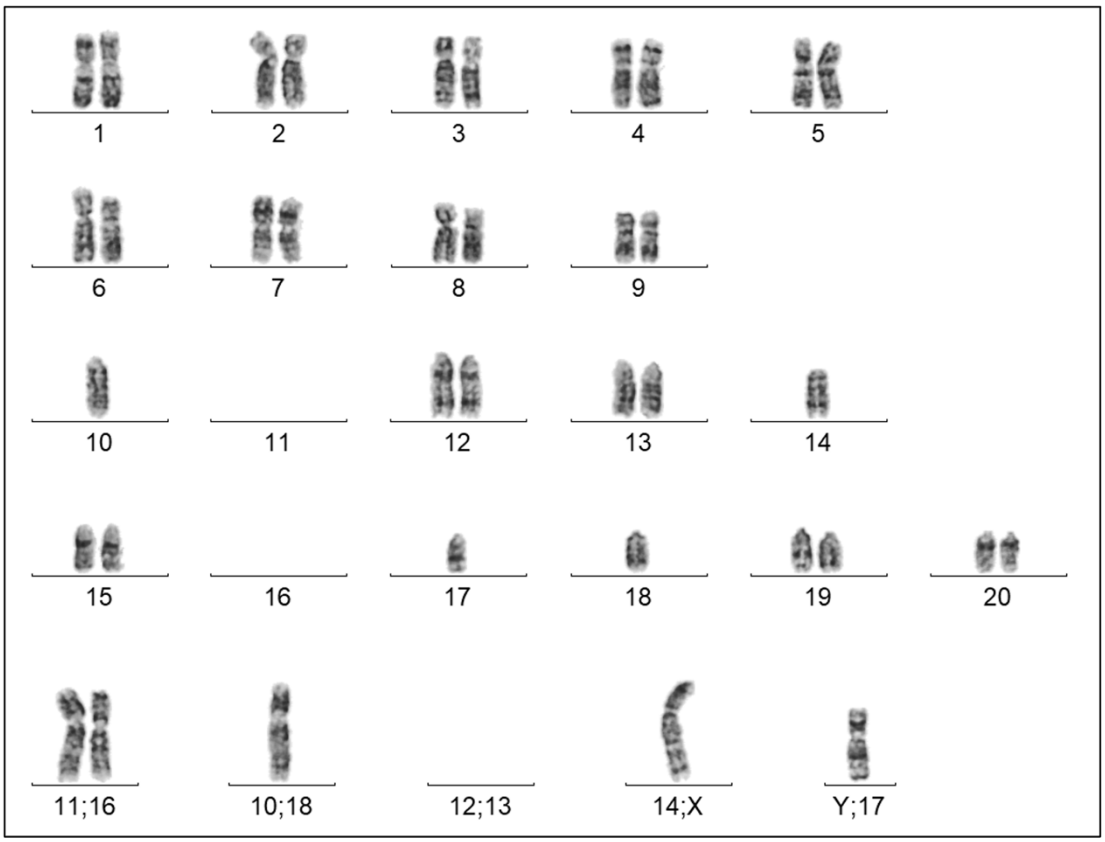

Figure 14. Reconstructed karyotype of the male *Nanger soemmerringii* cell line (NSO3). The chromosomal arrangement shows a diploid number of 2n = 37. The karyotype is characterized by a homozygous Robertsonian translocation between chromosomes 11 and 16 (11/16 +/+) and a polymorphic heterozygous fusion between chromosomes 10 and 18 (10/18 +/-), the latter resulting in the presence of both a metacentric element and its acrocentric counterparts. Sex chromosomes are identified as XY.

AIMS OF THE WORK

The centromere is a nucleoprotein structure essential for maintaining genomic integrity during cell division. While canonical centromeres are typically associated with large arrays of satellite DNA, "neocentromeres" can arise at novel chromosomal loci devoid of such repeats (satellite-free). The genus *Equus* (order Perissodactyla) has historically served as the primary model system for studying satellite-free centromeres. Recently, other centromeres devoid of satellite DNA were characterized in other Perissodactyla species (*Tapirus indicus*). However, the extent of this phenomenon in other mammalian orders, such as Artiodactyla, remains largely unexplored.

This work focuses on two gazelle species, *Nanger dama* and *Nanger soemmerringii*, which are characterized by high genomic instability and plastic karyotypes shaped by frequent chromosomal rearrangements.

The primary aim of this thesis is to investigate, through the integration of cytogenetic and molecular techniques the centromeric landscape, satellite DNA organization and the distribution of telomeric repeats within *N. dama* and *N. soemmerringii*. Specifically, we want to verify whether the high genomic instability that characterizes the karyotypes of these species led to the formation of evolutionary recent satellite-free centromeres in *Nanger* species.

MATERIALS AND METHODS

1. Cell culture

In this study, primary fibroblast cell lines from *Nanger dama* (derived from an individual housed at from the San Antonio Zoo, Texas, USA) and *Nanger soemmerringii* (Saint Louis Zoo, Missouri, USA) were used. Both cell lines were kindly provided by Terje Raudsepp (Texas A&M University, USA). Primary fibroblasts were cultured in high glucose Dulbecco's Modified Eagle Medium (DMEM), supplemented with 20% fetal bovine serum, 2 mM L-glutamine, 1% penicillin/streptomycin and 2% non-essential amino acids. Cells were maintained in a humidified atmosphere of 5% CO_2 at 37°C.

2. Identification and annotation of satellite repeats

Satellite DNA families in the two gazelle species were identified by TAREAN (TAndem REpeat Analyzer) analysis, a computational tool designed to detect satellite repeats directly from unassembled short reads (Novák *et al.,* 2017), starting from genomic reads derived from two distinct ChIP-seq (Chromatin ImmunoPrecipitation and Sequencing) experiments conducted with CREST (Calcinosis, Raynaud's phenomenon, Esophageal dysmotility, Sclerodactyly, and Telangiectasia) serum on primary fibroblasts from *Nanger dama* and *N. soemmerringii*, respectively. Further details regarding the ChIP-seq experimental procedures will be provided in paragraph 6. For both species, TAREAN analyses were performed using both ChIP and Input (non-immunoprecipitated control) reads. The identified satellite repeats were mapped on *Nanger dama* genome assembly (paragraph 8) using RepeatMasker (Galaxy Version 4.0.9) available at the Galaxy web platform (Community 2022).

3. Satellite DNA isolation

To amplify by PCR (Polymerase Chain Reaction) the genomic loci containing satellite repeats within *Nanger dama* genome, specific primers were designed (**Table 1**) based on the consensus sequence of the major satellite DNA (NangerSat1) identified by TAREAN in the genus *Nanger*.

In order to obtain several copies of amplified satellite DNA, 15 identical PCR reactions were performed in parallel. The specific composition of a single reaction and the thermal cycling profile are detailed in **Tables 2** and **3**, respectively. Both 5X Green GoTaq® Reaction Buffer and GoTaq® DNA Polymerase were provided by Promega Corporation (Madison, WI, USA). To ensure the absence of DNA contamination, a negative control (No-Template Control, NTC) was included, in

which genomic DNA was replaced with an equivalent volume of nuclease-free water. PCR products were loaded onto a 1% agarose and separated by electrophoresis at 40 V for approximately one hour. To determine the size (bp) of the resulting DNA fragments, a molecular weight marker (GeneRuler™ 1 kb DNA Ladder, Thermo Scientific™) was loaded alongside the samples.

Primer forward	**5' CTGGCTAGGAGAGAGCTAGGGAA 3'**	**23 bp**	**Melting temperature 65.90 °C**	**GC% 56.52**	**Secondary structure moderate**
Primer reverse	**5' GGTCTGCCTCTGAGTGTCTGG 3'**	**22 bp**	**Melting temperature 66.10°C**	**GC% 59.09**	**Secondary structure absent**

Table 1. Sequences and general features of the forward (F) and reverse (R) primers used to amplify satellite DNA in *Nanger dama* genome.

Reaction mix	**Volumes**
H2O *Nuclease-free*	15,75 µl
Promega Buffer 5X Green GoTaq™	5 µl (1X)
dNTPs (10 mM)	5 µl (0,2 mM)
Primer Forward (100µM)	0,15 µl (0,6 µM)
Primer Reverse (100µM)	0,15 µl (0,6 µM)
GoTaq® DNA polymerase (5 U/µl)	0,2 µl (1U)
Genomic DNA (100 ng/µl)	2 µl (200 ng)
Final volume 25 µl	

Table 2. PCR reaction

95°C x 2 min e 30 sec	
94 °C x 1 min 67°C x 1 min 72°C x 3 min	3 cycles
94°C x 40 sec 67°C x 40 sec 72°C x 3 min	32 cycles
72°C x 4 min e 55 sec	
Hold: 10°C	

Table 3. PCR profile

To isolate the amplified satellite sequences, target DNA fragments were extracted and purified from the agarose gel following electrophoretic separation. This procedure was performed using the Wizard® SV Gel and PCR Clean-Up System (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions. The bands of interest were excised from the gel using a sterile scalpel. The resulting gel slices were treated with Membrane Binding Solution (10 µL per 10 mg of gel) and incubated at 60°C for approximately 10 minutes (until the agarose was completely solubilized). Each sample was then transferred to a spin column assembly and incubated at room temperature for 4 minutes to allow for optimal DNA binding. The columns were centrifuged at 16,000 g for 2 minutes, and the flowthrough was discarded. Two subsequent washing steps were performed

using Membrane Wash Solution, followed by centrifugation at 16,000 g for 2 minutes. Finally, the bound DNA was eluted in 40 µl of nuclease-free water by centrifugation at 16,000 g for 1 minute. To confirm the successful extraction and integrity of the purified fragments, the eluted DNA was loaded and visualized on a 1% agarose gel.

4. Fluorescence *in situ* hybridization (FISH)

4.1. Metaphase chromosome preparation

Mitoses from *Nanger dama* and *Nanger soemmerringii* were mechanically detached from the plate by "blowing" the culture medium on the dish surface (Piras *et al.,* 2010). The cell suspension was centrifuged and incubated in a hypotonic solution (0.075 M KCl) at 37°C for 20 minutes. Subsequently, cells were fixed in cold (-20°C) methanol: acetic acid (3:1) solution. The resulting chromosome preparations were spread onto glass slides.

Slides were aged in a dry oven at 90°C for 90 minutes and treated with pepsin (50 mg/mL) in 0.01 M HCl. Following enzymatic digestion, slides underwent a series of washes: 1X PBS for 2 minutes; a solution containing 1X PBS and 0.05 M MgCl2 for 5 minutes; a fixing solution (1X PBS, 0.05 M MgCl2, and 4% PFA) for 5 minutes; a final wash in 1X PBS for 5 minutes. The slides were then dehydrated through an increasing ethanol series (70%, 90%, and 100%).

4.2. Probe labeling and hybridization

Two distinct FISH experiments were performed using either a satellite DNA probe or a telomeric oligonucleotide probe on metaphase spreads of both *Nanger dama* and *N. soemmerringii*.

- Satellite Probe: the 2,200 bp PCR-amplified satellite DNA fragment was utilized.
- Telomeric Probe: A mixture of synthetic fragments (1–20 kb) was used.

Both probes were labelled by nick-translation with Cy3-dUTP at 15°C (15 minutes for the satellite probe and 20 minutes for the telomeric probe). The labelling reaction was stopped with 4 µl of 0.5M EDTA. Probes were precipitated with salmon sperm DNA (0.05 mg/µl final concentration), sodium acetate (pH 4, 0.09 M), and 100% ethanol. After incubation at -20°C for 2 hours and centrifugation at 13,000 rpm for 20 minutes, the pellets were washed in 70% ethanol, dried at 37°C, and resuspended in hybridization solution (10% dextran sulfate, 2X SSC and 1% TWEEN).

The resuspension volume and formamide concentration in hybridization solution were adjusted according to the required stringency. Satellite DNA probe was resuspended in 60 µl of hybridization solution (50% formamide) and 25 µL of probe were put on slide. Differently, telomeric probe was resuspended in 40 µL of hybridization solution (25% formamide) and hybridization was performed using 20 µL of probe per slide.

4.3. Denaturation and post-hybridization washes

Both the probe and the metaphase spread preparation were simultaneously denatured at 72°C for 3 minutes and 30 seconds. Hybridization was performed overnight at 37°C in a humid chamber.

Post-hybridization washes were conducted to achieve specific stringency levels (83% and 57%, respectively):

- Satellite DNA probe: three 5-minutes washes at 51°C in 0.1X SSC, followed by a 5-minutes wash in 2X SSC at room temperature.
- Telomeric DNA probe: three 5-minute washes at 37°C in 25% formamide/4X SSC, followed by three 5-minute washes in 4X SSC at 37°C.

5. Immunofluorescence (IF)

In order to perform ChIP-seq experiments with CREST (Calcinosis, Raynaud's phenomenon, Esophageal dysmotility, Sclerodactyly, and Telangiectasia) serum on chromatins extracted from *Nanger dama* and *N. soemmerringii* primary fibroblasts. The specificity of the CREST serum for the gazelle CENP-A protein was validated via immunofluorescence analysis on *N. dama* metaphase spreads.

Metaphase chromosome preparation was obtained as explained in paragraph 4.1. Approximately 100 µL of the cell suspension were cytosun (Cytospin BHG Hermle Z380) at 1,250 rpm for 8 minutes, creating two separate spots. Slides were fixed in ice-cold methanol for 4 minutes and subsequently incubated in 1X PBS containing 0.05% Tween-20 (PBST). For primary labelling, the slides were incubated with human CREST serum (kindly provided by Dr. Claudia Alpini, IRCCS Policlinico San Matteo, Pavia, Italy), diluted 1:250 in PBST, for one hour at 37°C. Following three 5-minute washes in PBST at room temperature, the slides were incubated with an Alexa Fluor 488-conjugated anti-human secondary antibody (diluted 1:100 in PBST) for one hour at 37°C. After three additional washes in PBST, chromosomes were counterstained.

6. Visualization and image analysis

To visualize FISH and IF analyses, chromosomes were counterstained with 0.2 µg/ml DAPI (4′,6-diamidino-2-phenylindole) and slides were mounted using McIlvaine mounting medium. Grayscale images were captured using a Zeiss Axio Scope.A1 fluorescence microscope equipped with a CCD Photometrics camera. Image pseudo-coloring and merging were performed using IpLab software.

7. Chromatin ImmunoPrecipitation and sequencing (ChIP-seq)

Two independent Chromatin Immunoprecipitation experiments followed by high-throughput sequencing (ChIP-seq) were performed on chromatins extracted from primary fibroblasts of *Nanger dama* and *Nanger soemmerringii*. Chromatin immunoprecipitation (IP) allows the identification and characterization of DNA-protein interactions through the use of a specific antibody directed against the protein of interest. For each immunoprecipitation (IP) reaction, at least 10 million cells were harvested and centrifuged at 1,700 rpm for 7 minutes. To fix the protein-DNA interactions, *cross-linking* was performed by adding 1% formaldehyde and incubating the cells under agitation for 15 minutes at 26°C. The reaction was blocked by the addition of 0.125 M glycine for 10 minutes at 26°C. The suspension was centrifuged at 800 rcf for 5 minutes at 4°C, and the resulting pellet was stored at -80°C overnight. The pellet was subsequently thawed on ice and washed twice with 1X PBS supplemented with protease inhibitors, followed by centrifugation at 800 rcf for 6 minutes at 4°C. The final pellet was resuspended in ChIP lysis buffer (0.25% SDS, 50 mM Tris-HCl pH 8, 10 mM EDTA pH 8) containing a Protease Inhibitor Complex (PIC). Samples were divided into aliquots of about 20 million cells per 650 µl. The chromatin was then sonicated (Branson Sonifier 250) to generate DNA fragments ranging from 200 to 800 bp. The size was verified by agarose gel electrophoresis. Following fragmentation, the samples were centrifuged at 13,000 rpm for 10 minutes at 4°C, and the supernatant was brought to a final volume of 6400 µl with dilution buffer (0.5% Nonidet P40, 10 mM Tris-HCl pH 7.5, 2.5 mM MgCl2, 150 mM NaCl) supplemented with PIC. The pre-clearing phase followed, in which agarose beads were used to block sites that might form non-specific bonds with the antibody in subsequent steps. Pre-clearing was performed using Protein G Sepharose™ 4 Fast Flow beads (GE Healthcare) under agitation at 4°C for one hour, followed by centrifugation at 4,000 rpm for 5 minutes at 4°C. From the resulting supernatant, 120 µl was reserved as the Input (non-immunoprecipitated control). The remaining volume was divided into aliquots and incubated overnight at 4°C with the human CREST serum. Subsequently, the immunocomplexes were captured by incubation with Protein G beads for 3 hours at 4°C under agitation. The beads were recovered by centrifugation at 1200 g for 2 minutes at 4°C, and the supernatant was removed. The pellet was washed 5 times with ChIP wash buffer (0.25% SDS, 1% Triton X-100, 2 mM EDTA pH 8, 150 mM NaCl, 20 mM Tris-HCl pH 8) and once with ChIP final wash buffer (containing 500 mM NaCl). After removing the final wash, immunocomplexes were eluted in ChIP elution buffer (1% SDS, 100 mM NaHCO3, 40 µg/mL RNase A) for 15 minutes at room temperature, followed by one hour at 37°C. Reverse cross-linking was performed by incubating the samples at 65°C overnight.

The eluted DNA was purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) and quantified via the Quantus™ Fluorometer (Promega). The immunoprecipitated DNA fragments were

subjected to paired-end sequencing on the Illumina NovaSeq 6000 platforms at IGA Technology Services (Udine, Italy).

8. Bioinformatic analysis of ChIP-seq data

The Ruminant Telomere-to-Telomere (RT2T) Consortium (Kalbfleisch *et al.,* 2024) provided us a preliminary assembly of *Nanger dama* genome, with most chromosomes represented by a single contig. The ChIP-seq reads of *Nanger dama* were aligned on the RT2T genome using Bowtie2 with the default parameters (version 2.4.2) (Langmead *et al.,* 2009; Langmead & Salzberg, 2012). In the same way, the ChIP-seq reads of *Nanger soemmerringii* were aligned on the same genomic assembly. The normalization of the ratio between ChIP reads and input reads was performed using bamCompare, available in the deepTools suite (version 3.5.0) (Ramírez *et al.,* 2016). The alignment files were visualized using IGV (Integrative Genomics Viewer) software (version 2.9.2). Images of enrichment peaks and satellite annotation were obtained using pyGenomeTracks tool (3.6 version) (Lopez-Delisle *et al.,* 2021).

9. Comparative genomic analysis

A whole-genome alignment between *Nanger dama* genome assembly and *Bos taurus* reference genome (NCBI *Bos taurus* ARS-UCD1.2, GenBank assembly GCA_002263795.4) was performed using Chromeister (Pérez-Wohlfeil *et al.,* 2019). The analysis was conducted via the Galaxy EU platform (Chromeister ultra-fast pairwise genome comparisons) (Afgan *et al.,* 2016). Both genomes were provided in FASTA format, and the alignment was performed utilizing default parameters.

10. Identification of telomeric repeats

Telomeric repeats distribution in *Nanger dama* assembly was obtained by BLASTN 2.11.0 search for the $(TTAGGG)_4$ motif using the parameters -evalue 10 -word_size 11 -gapopen 5 -gapextend 2 reward 2 -penalty -3 -outfmt 7. Hits within 250 bp of each other were merged into single loci using Bedtools v2.30.0.

RESULTS

1. Identification and amplification of the major centromeric satellite in the genus *Nanger*

To characterize satellite DNA in the gazelle species *Nanger dama* and *N. soemmerringii*, we performed an analysis using TAREAN (TAndem REpeat ANalyzer), a graph-based computational tool designed to detect satellite repeats directly from unassembled genomic short reads (Novàk *et al.,* 2017). The genomic reads used for satellite DNA detection derive from two distinct ChIP-seq (Chromatin ImmunoPrecipitation sequencing) experiments conducted with CREST (Calcinosis, Raynaud's phenomenon, Esophageal dysmotility, Sclerodactyly, and Telangiectasia) serum on primary fibroblasts from *Nanger dama* and *N. soemmerringii*, respectively.

TAREAN provides monomer consensus sequence and relative genomic abundance for each satellite DNA family. All detected satellites are reported in **Table 4**, where satellites sharing high sequence identity are indicated with the same colour. According to the evolutionary proximity between the two gazelle species, we identified a common major satellite (green in Table 4), whose 700 bp-long consensus sequence is reported in **Figure 15**. This satellite (which we named **NangerSat1**) resulted enriched in ChIP reads in both species, since the normalized ratio between ChIP and Input reads aligned on satellite sequence is 2.31 (Table 4). Thus, we can infer that this satellite is centromeric. Interestingly, we found high sequence identity between NangerSat1 and the satellites OSSAT1 and BTSAT6 (green in Table 4), from the Artiodactyls *Ovis aries* and *Bos taurus*, respectively.

As reported in **Table 4**, we identified a 785 bp-long satellite exclusively in *Nanger dama* reads (blue in Table 4), which we named **NangerSat2**. This satellite (monomer reported in **Figure 16**) shares high sequence identity with other satellites of *Ovis aries* and *Bos taurus* (OOSAT2 and BTSAT4, respectively). NangerSat2 did not result enriched in *Nanger dama* ChIP reads.

Finally, TAREAN detected another minor 732 bp-long satellite (yellow in Table 4) in *Nanger soemmerringii*, which does not show sequence conservation with known Artiodactyls satellites and is not enriched in ChIP reads.

Species	Type of reads	Satellite name by TAREAN	Consensus length (bp)	Proportion (%)	Normalized ratio ChIP/Input	Artiodactyla satellites with sequence identity
Nanger dama	Input	CL9	788	0.46	1.39	OSSAT2; BTSAT4
		CL10	701	0.41	2.31	OSSAT1; BTSAT6
	ChIP	CL7	699	0.9	2.31	OSSAT1; BTSAT6
		CL9	785	0.66	1.39	OSSAT2; BTSAT4
Nanger soemmerringii	Input	CL9	702	0.36	2.26	OSSAT1; BTSAT6
	ChIP	CL7	702	0.8	2.26	OSSAT1; BTSAT6
		CL1	732	12	1.02	

Table 4. **Satellite DNA families identified by TAREAN analysis in two gazelle species**. All the satellites detected using ChIP and Input reads are reported with the following features: satellite name given by the tool, the length of the consensus sequence, the percentage of reads containing the satellite out of the total of reads analysed, the normalized ratio between ChIP and input reads (calculated using Bowtie2 from Galaxy platform), and the orthologous Artiodactyla satellites. Satellites exhibiting high sequence identity are indicated with the same colour.

```
>NangerSat1
TTCCCTCTGACAGACCTCTCACTCCTGGAGCCCTTGGCAGTGTGGAAGCCACCAGCCCCTAACCCCTCCCCAAGGAGGCCT
GGGGGCCTAGGTCGCAAGGCAAGGCAGGAGCGGAGCTGAGGCCAGGAAGGAGCCAGTGCTGTCCCCCGGGCGCCCACAGCC
TGCACCTTCCCGGGCAGGAAGCCTAGCAGCTAGGCAAGTCTGGGCAAGGCTCGCCCAGGGCAGAGCAGCCTCGTGCCGGGG
GCTCAGCCAGATGGGCCTCTGAGCAGAGCTGGAGCTCTGCCGTGCTGGAGGCACACTCGGGCCTTCCCCATCAGGGCTTGC
AGATGCACGCAGCGGGGCACCCTCCTCCTCCCCAAGAGGGTCGCTTCCCCGCATGCTGTGCACTCGGCTTCCAGGCGGGGT
CGCCAGCGTGCCTTCTGGAGGTCTGGGCTCAGGGGCTGGTTCCCTGAGCTGGGGCTGTGGGCGGGAGCCCAGACACTCAGA
GGCAGACCCAAATCACCTGCACACGTCTCTGCTGGCTAGGAGAGCACTAGGAACCCACGTCTGTCAAGCCCTTGCCGGGGG
TGAACGGAGCCCCACTCACCGGTGTGGCCCTGCACACCCCACCCAGGGGCCCTGGAAGCCTGGTGCTCTCCGTGCGGCGCT
GGGTCTGTGCAGCCAAGCAACCACTGGAGAAACTATCCAGTTGCCCTAGCC
```

Figure 15. Consensus sequence of NangerSat1 (701 bp-long)

```
>NangerSat2
CGTCTCGGTGCTCAGTCCTCGACTTGTGCTGAGGCCGGAGCCTGAGCTTTCCTCGACCGTCCTGACCTAGGTCTTGGGGTT
CCCCTGAGGCCCCCACAGAGGAGACAGGCCTCGTCTCGCCTGGACACGTGCACGGCTACTTGCCTGACCAGCTGCAGCAGC
AGTGTCAGGCTCCATGTCGAGCGGACACAGGGATCTGTGGCTTGCCCTCGAGGTCCAAAGTGCTGTCACACGTGCCAGTGT
GTCGGGCGTCGAGCCTCGGGGTGACAGTCAAGGCAGTGCAGGGCAGTCAGGTTCATCTGGAGTGGACAAGGCATTTGGGGG
TCTTGGGGAATGGTGGCAAGACCCCTGATGCTCCTCCCGACTGTCCTGGTGCGAGTGTCTCCTGTTGGGATGTGAGGGAAG
CGCTGGGAACCCTTTCCCGACAGAGCAGGAAAGGGACCCTCATCTGGAGCTTAGCAGGGGAAAACGGGGCTCCTCTTGAGT
CGCAGCGGGTCCCTTGGTGTTCCTCTCGAGTGGGGACGGGTGGGGCGGGGAACTTCTTGGGTCACAGCAAGGGGGTGAAGG
AGCCTTTCGAGGTCCAAGAGGGAAGGGGTGATTTCCCCCGAGACGCCACAGCGGATAAGGGCCTCACCCCGCTTGCAAGGG
AGAAGCTTCTGGTTTTTCTCGAGTTGTGGCGGGGTTTCTCGAGTTACGACGGGGAGCCCATGGACCCGCCCCTGTGGCCTC
AGGAGAGGCCAGTCTCCGTGCCAGTTGCCAGGCGCCTCTGGGGATTCTTCCCCAAT
```

Figure 16. Consensus sequence of NangerSat2 (785 bp-long)

In order to amplify the genomic loci containing NangerSat1 repeats in *Nanger dama* genome, PCR analysis was performed using the primer pair (previously) described in the Material and Methods section. The resulting PCR products were visualized by agarose gel electrophoresis (**Figure 2A**), together with a negative control (CT), which confirmed the absence of contamination or primer-dimer formation, and a molecular weight marker (1kb ladder). Four different DNA fragments were generated by PCR, 690 bp, 780 bp, 1460 bp and 2210 bp long, respectively. Given that the NangerSat1 monomer is 700 bp long, the first two fragments (690 bp and 780 bp) correspond to approximately one satellite repeat, while the third fragment (1460 bp) represents two repeats. Finally, the fourth fragment (2210 bp) corresponds to three tandem repeats of NangerSat1 satellite. The differences in DNA fragment lengths result from the variable primer annealing sites across satellite arrays in *Nanger dama* genome. These four DNA fragments were visualized as discrete bands (**Figure 17A**), then excised and purified from the agarose gel. To verify the purification of the DNA fragments, the four eluted samples were subsequently analysed by a second agarose gel electrophoresis (**Figure 17B**). The longest (2210 bp) purified fragment, highlighted in orange, was selected as a probe for the two FISH experiments on the metaphase spreads of *Nanger* species described in the following section. By incorporating multiple repeat units, this longer probe is expected to provide higher hybridization efficiency with chromosomal DNA.

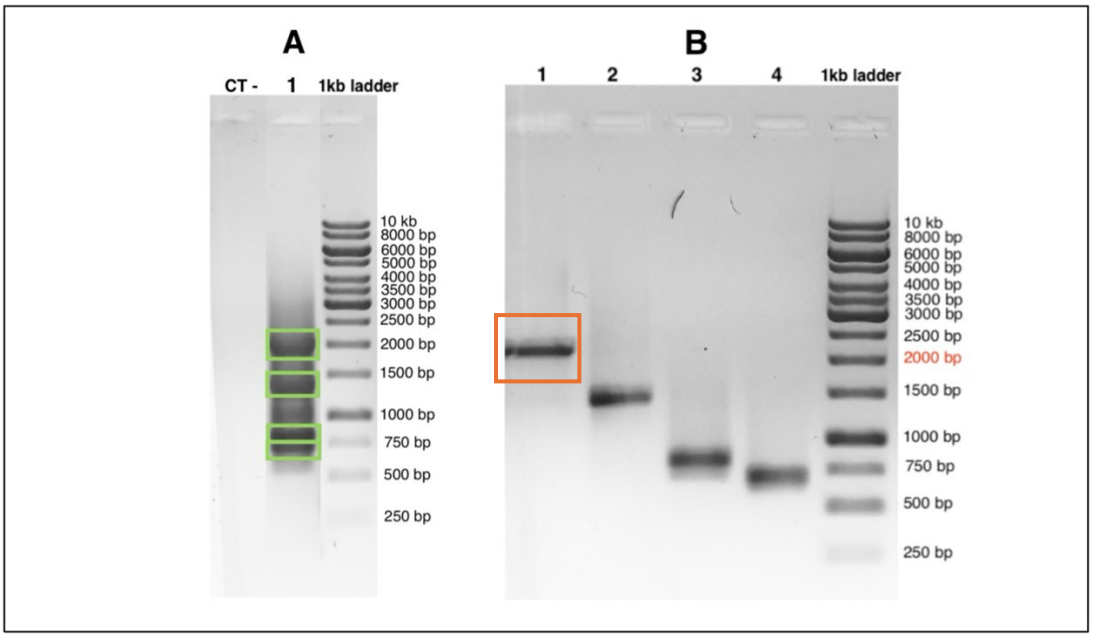

Figure 17. Electrophoretic analysis of satellite DNA amplification products. (A) Agarose gel electrophoresis showing, from left to right: negative control, PCR products, and molecular weight marker (1kb DNA ladder). The four bands highlighted in green represent the DNA fragments subsequently excised and purified. (B) Analysis of the four purified DNA fragments loaded into separate wells. The fragment highlighted in orange (corresponding to 2210 bp) was selected as a probe for the FISH experiment.

2. Cytogenetic identification of chromosomes with satellite-free centromeres in *Nanger dama* and *N. soemmerringii*

The 2210 bp fragment previously obtained by PCR was used as a probe in two FISH experiments performed on metaphase chromosomes of *Nanger dama* (**Figure 18**) and *N. soemmerringii* (**Figure 19**), respectively. We performed both FISH experiments with the same satellite probe, since we previously demonstrated that these two species share the same satellite DNA family, NangerSat1 (**Table 4** and **Figure 15**).

In both FISH experiments, DAPI counterstain (blue) was used to visualize the metaphase chromosomes and reveal the banding patterns, while the NangerSat1 fluorescent probe (red) indicates the chromosomal localization of satellite DNA.

As shown in **Figure 18**, hybridization signals are present at the primary constrictions of the majority of *N. dama* chromosomes (36 out of 39), suggesting that they contain satellite-based centromeres. Conversely, the centromeres of three chromosomes (two submetacentric and one acrocentric) lack hybridization signals, suggesting that these loci may be satellite-free.

Similarly, as shown in **Figure 19**, hybridization signals were detected in the majority (34 out of 37) of centromeres in *N. soemmerringii*, suggesting that they are satellite-based. Three chromosomes (two submetacentric and one acrocentric) are devoid of FISH signals at their primary constrictions, suggesting that they may correspond to satellite-free centromeres.

FISH experiments confirm that NangerSat1 is the major centromeric satellite in both *Nanger* species, as suggested by *in silico* TAREAN analysis. More interestingly, we obtained the first cytogenetic evidence that both species harbor three centromeres devoid of satellite DNA.

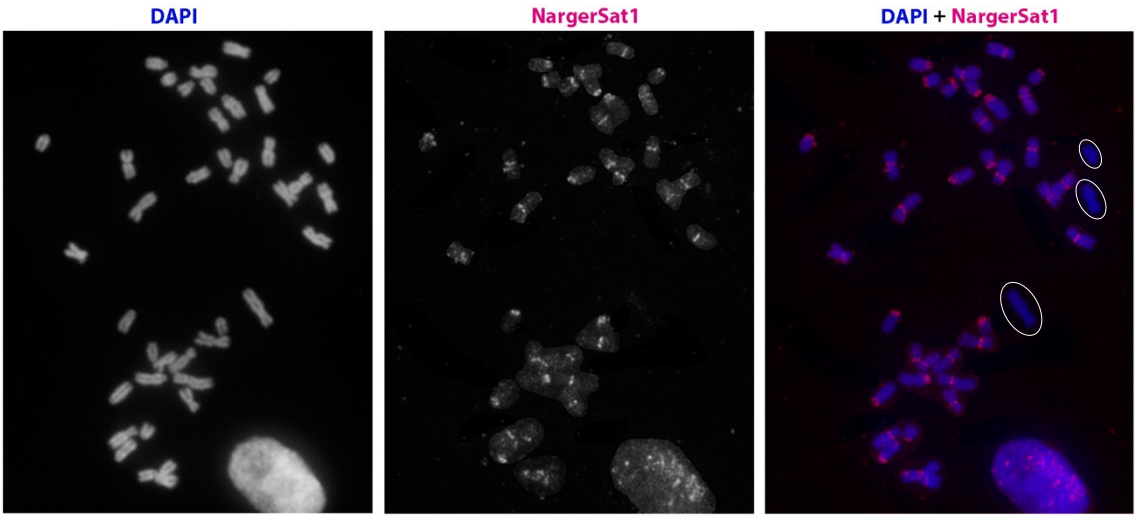

Figure 18. **Localization of satellite DNA on metaphase chromosomes of *Nanger dama***. FISH signals (in red) were obtained using a DNA fragment containing three NangerSat1 repeats as a probe. The metaphase chromosomes were counterstained with DAPI (in blue). The chromosomes whose primary constrictions are devoid of hybridization signals are marked by white circles.

Figure 19. **Localization of satellite DNA on metaphase chromosomes of *Nanger soemmerringii***. FISH signals (in red) were obtained using a DNA fragment containing three NangerSat1 repeats as a probe. The metaphase chromosomes were counterstained with DAPI (in blue). The chromosomes whose primary constrictions are devoid of hybridization signals are marked by white circles.

3. Molecular characterization of centromeres in two *Nanger* species by ChIP-seq analysis

To characterize the organization of centromeres in the two *Nanger* species at the molecular level, we planned to perform two distinct ChIP-seq experiments on chromatin extracted from the primary fibroblasts of both species. Since a species-specific antibody against gazelle CENP-A, the epigenetic determinant of centromeric function, is not available, chromatin from both species was immunoprecipitated using a human CREST serum. This serum, derived from patients diagnosed with CREST syndrome (Calcinosis, Raynaud's phenomenon, Esophageal dysmotility, Sclerodactyly, and Telangiectasia), is usually suitable for the immunoprecipitation of centromeric DNA sequences. Indeed, it is enriched in various antibodies directed against centromeric proteins, including the histone variant CENP-A. Consequently, CREST serum can effectively substitute for a specific anti-CENP-A antibody in identifying centromeric domains (Nergadze *et al.,* 2018).

Initially, we validated the CREST serum by immunofluorescence analysis on *Nanger dama* metaphase spreads. As shown in **Figure 20**, all gazelle primary constrictions are bound by the CREST serum: each primary constriction is labelled by two green spots, corresponding to the centromeres of the two sister chromatids. The equivalent validation was not conducted on *N. soemmerringii* metaphase chromosomes, considering the close phylogenetic relationship between the two *Nanger* species under study.

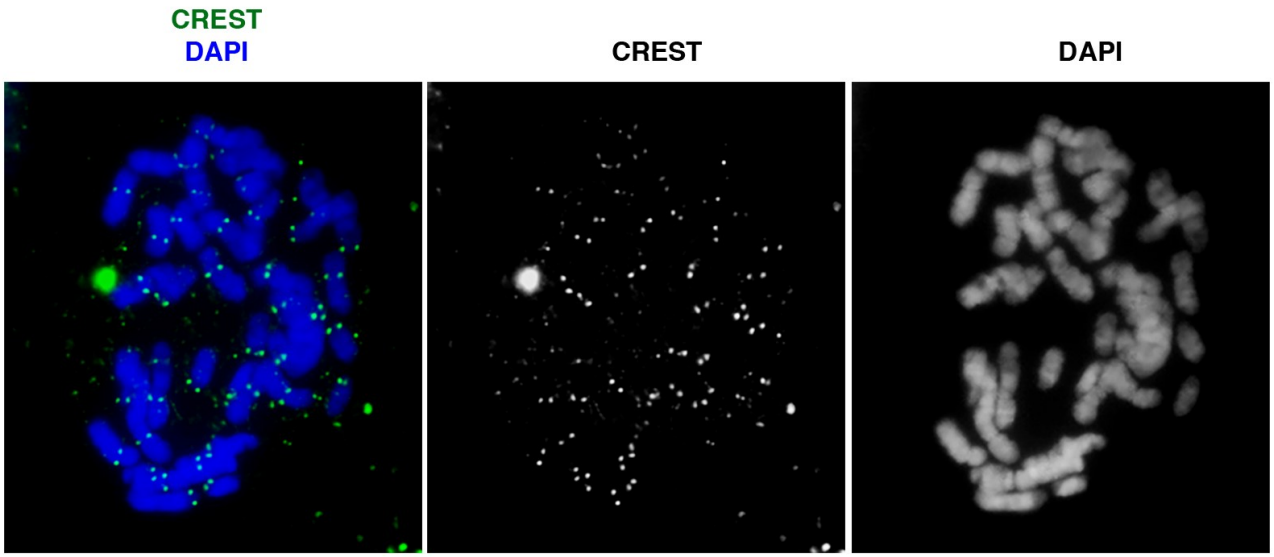

Figure 20. **Immuofluorescence with the CREST serum on *Nanger dama* metaphase chromosomes**. The localization of human serum CREST (green spots) can be observed in all centromeres of metaphase chromosomes counterstained with DAPI (in blue).

After immunofluorescence validation, we carried out two ChIP-seq experiments with the CREST serum on chromatin extracted from primary skin fibroblasts of *Nanger dama* and *N. soemmerringii*.
The Ruminant Telomere-to-Telomere (RT2T) Consortium (Kalbfleisch *et al.,* 2024) provided us a preliminary assembly of *Nanger dama* genome, containing most chromosomes represented by a single contig. Specifically, the diploid assembly includes two haplotypes (named Dama_gazelle _primary and Dama_gazelle_secondary). These two haplotypes do not correspond to maternal and paternal sets, rather, the "primary" haplotype contains the more complete and better-assembled chromosomes, while the "secondary" haplotype includes those with lower assembly quality.
The ChIP-seq reads from both species were mapped onto both *Nanger dama* haplotypes. However, due to the poor assembly quality of the secondary haplotype, we were unable to reliably identify all the centromeric domains of this haplotype, particularly those containing extended satellite DNA arrays. Consequently, in this work we will show only the centromeric domains of both species obtained by mapping ChIP-seq reads to the primary assembly.
The CREST enrichment peaks obtained from the alignment of *Nanger dama* ChIP-seq reads to the primary haplotype are reported in **Figure 21**, while **Figure 22** shows the enrichment peaks obtained from the alignment of *N. soemmerringii* ChIP-seq reads to the same primary haplotype.
In order to distinguish between satellite-based and satellite-free centromeres in both *Nanger* species, we localized the major 701 bp centromeric satellite (NangerSat1, colured in green) and the 785 bp minor satellite (NangerSat2, coloured in indigo). The satellite distribution is shown as coloured bars under the CREST enrichment peaks of both *Nanger* species (**Figure 21-22**). As expected, long arrays of NangerSat1(coloured in green) repeats underline the majority of gazelles' enrichment peaks, confirming it as the major centromeric satellite in this species. NangerSat1 arrays show size variation, ranging from 500 kb to 1.4 Mb.
As shown in **Figure 21**, we identified 13 centromeric domains containing satellite DNA arrays in *Nanger dama*, specifically in chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14 and 16. Moreover, we detected arrays of NangerSat2 satellite, typically placed in the pericentromeric regions of chromosomes 2, 4, 6, XY2 and XY4. These arrays exhibit length heterogeneity, spanning from 50 kb to 1.1 Mb. The satellite-based enrichments peaks show irregular shapes and the satellite DNA arrays below appear fragmented.
Interestingly, the centromeric loci of the chromosomes named XY1, XY2 and XY4 do not contain NangerSat1 satellite, suggesting that they are satellite-free (**Figure 21**). Their CREST enrichment peaks exhibit regular and Gaussian-like shapes, indicating a more accurate underlying sequence assembly due to the lack of satellite DNA. In XY1 and XY4 we detected two distinct enrichment peaks, which may correspond to separate epialleles (Purgato *et al.,* 2015). These two domains are

separated by approximately 130 kb in XY1 and 370 kb in XY4. We could not detect enrichment peaks or satellite arrays in the remaining chromosomes of the primary assembly (11, 12, 15, 17 and XY3).

The molecular organization of centromeres in *Nanger soemmerringii* is shown in **Figure 22**. Similar to *N. dama*, the satellite NangerSat1 is the major centromeric satellite, while NangerSat2 arrays were detected in the pericentromeric domains of chromosomes 3, 5, 7, 9, XY2 and XY4. The extension of these satellite arrays are consistent with those previously described for *Nanger dama*, as the reference genome is the same.

According to this analysis, *N. soemmerringii* chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14 and 16 carry satellite-based centromeres, that display irregular shapes and are completely covered by long arrays of NangerSat1.

Similarly to *N. dama*, chromosomes XY1, XY2 and XY4 possess centromeres completely devoid of NangerSat1, while long arrays of the pericentromeric satellite NangerSat2 flank the enrichment peaks of chromosomes XY2 and XY4 (**Figure 22**). The shapes of enrichment peaks of XY1, XY2 and XY4 are almost Gaussian-like. The centromere of chromosome XY1 exhibits two enrichment peaks, divided by a 130kb-long genomic region, which may also represent an instance of separate epialleles (Purgato *et al.,* 2015). No enrichment peaks or satellite arrays were detected in the remaining chromosomes (12, 15, 17 and XY3).

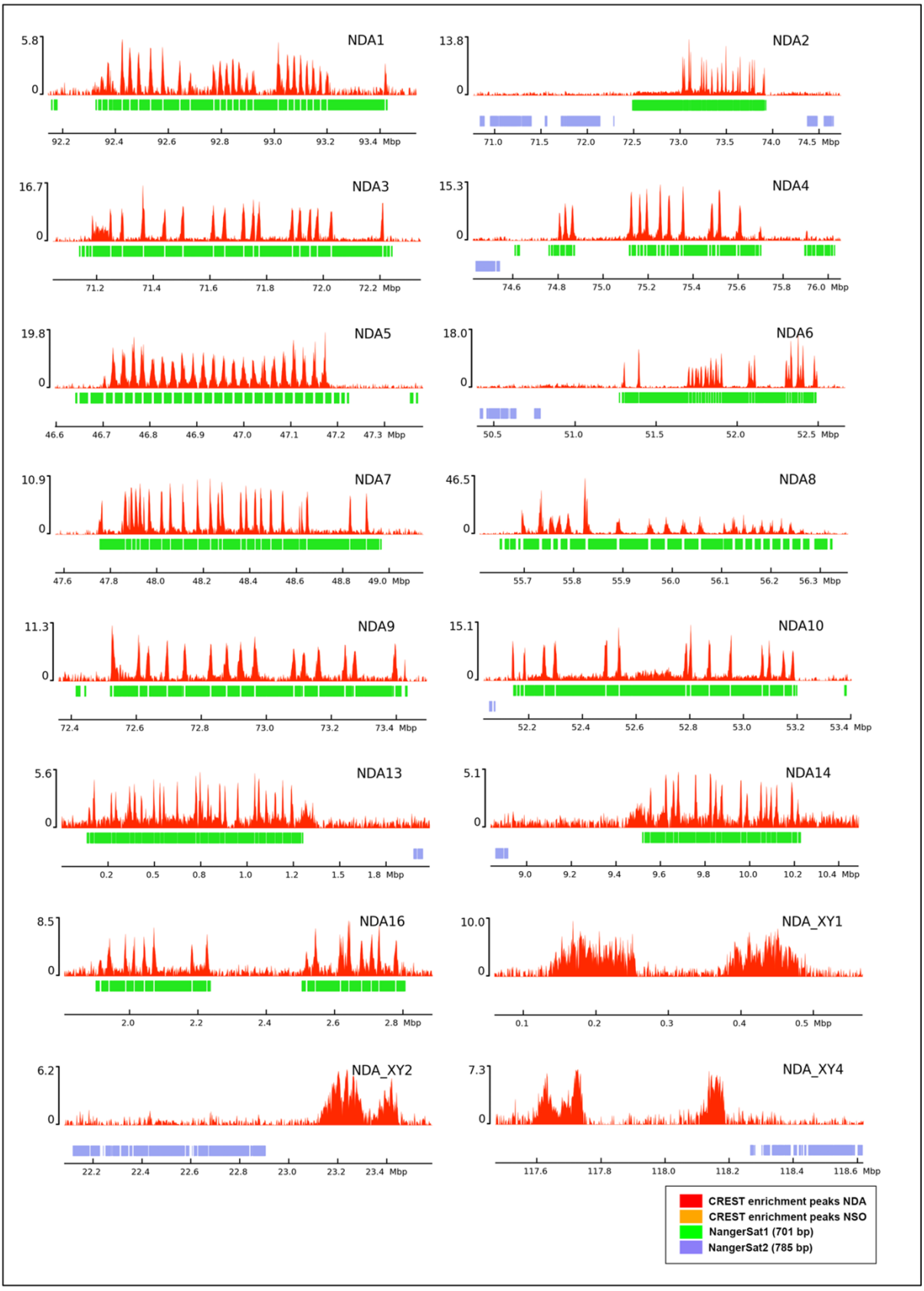

Figure 21. **Satellite-based and satellite-free centromeres in *Nanger dama***. For each centromere, the ChIP-seq profile of CREST serum is shown at the top. The *y*-axis reports the normalized read counts, whereas the *x*-axis reports the coordinates on gazelle genomic assembly. The colored bars represent the satellite DNA arrays.

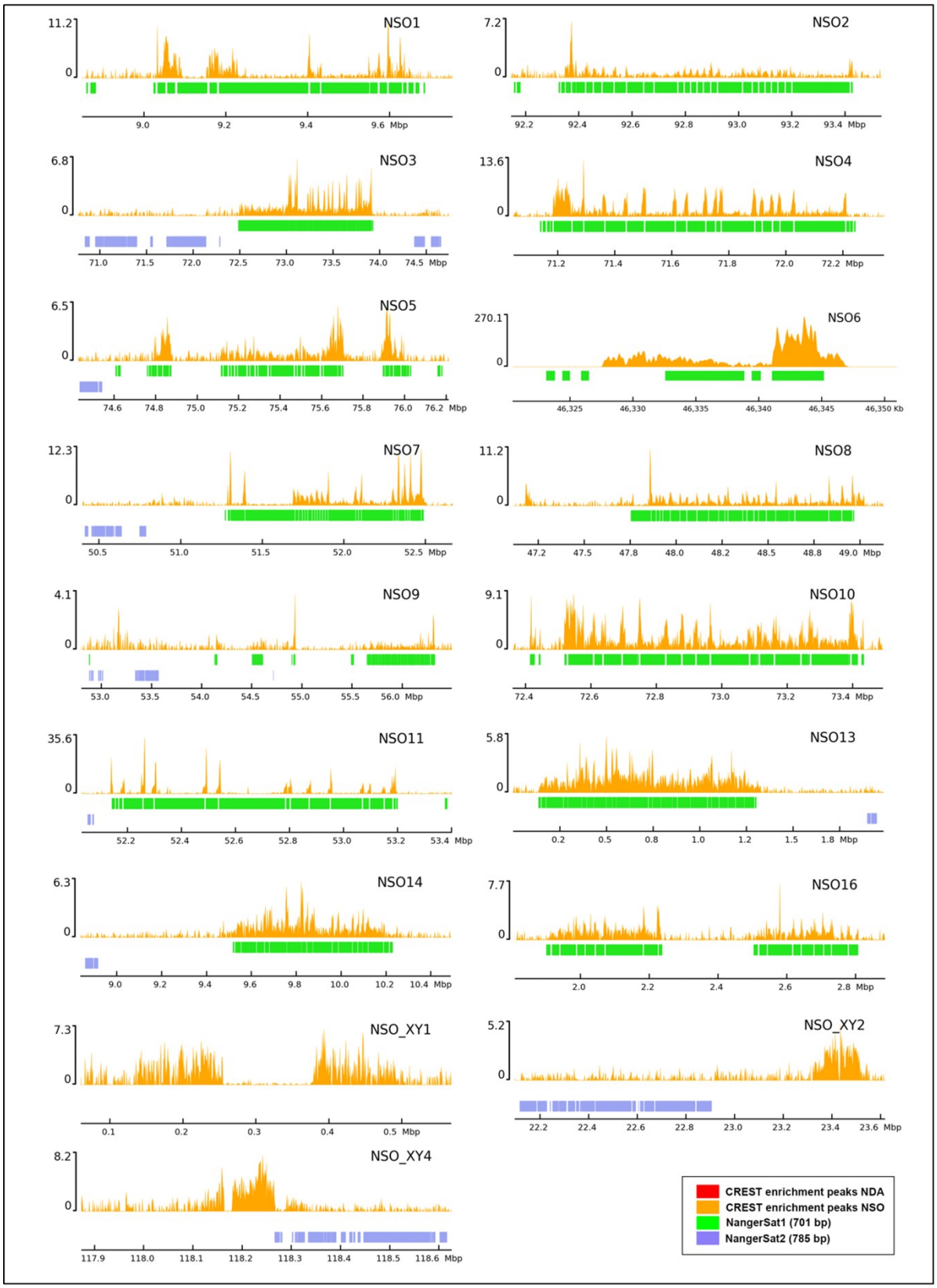

Figure 22. **Satellite-based and satellite-free centromeres in *Nanger soemmerringii.*** For each centromere, the ChIP-seq profile of CREST serum is shown at the top. The *y*-axis reports the normalized read counts whereas the *x*-axis reports the coordinates on gazelle genomic assembly. The colored bars represent the satellite DNA arrays (same colors used in figure 7).

To investigate whether the satellite-free centromeric organization is conserved between the two *Nanger* species under study, we compared CENP-A enrichment peaks positioning in chromosomes XY1, XY2, and XY4 (**Figure 23**). For all three examined chromosomes, the enrichment peaks (represented in red for *N. dama* and yellow for *N. soemmerringii*) are placed in orthologous positions in the two species.

In the XY1 chromosome of both *Nanger dama* (NDA) and *Nanger soemmerringii* (NSO), two distinct domains were detected. As shown in **Figure 23**, in NDA (red tracks), the first peak spans approximately 120 kb, while the second one covers roughly 110 kb. Similarly, in NSO (yellow tracks), we detected two distinct enrichment peaks. The first domain spans roughly 130 kb, followed by a second domain of approximately 120 kb. In both species, these two domains are separated by approximately 130 kb (measured from the end of the first peak to the start of the second).

Regarding the centromeric domains detected in XY2 chromosome (**Figure 23**), the CENP-A binding domain in *N. dama* exhibits an irregular profile, with a primary peak spanning approximately 180 kb, immediately followed by a narrower peak of roughly 100 kb. In contrast, there is a unique enrichment peak in *N. soemmerringii* in orthologous position, displaying a more regular profile and spanning approximately 160 kb.

Finally, the centromere of XY4 chromosome in *Nanger dama* comprises two distinct 370 kb distant peaks: the first spans approximately 150 kb, while the second roughly covers 100 kb. Differently, in *N. soemmerringii*, a single peak was detected, spanning approximately 140 kb.

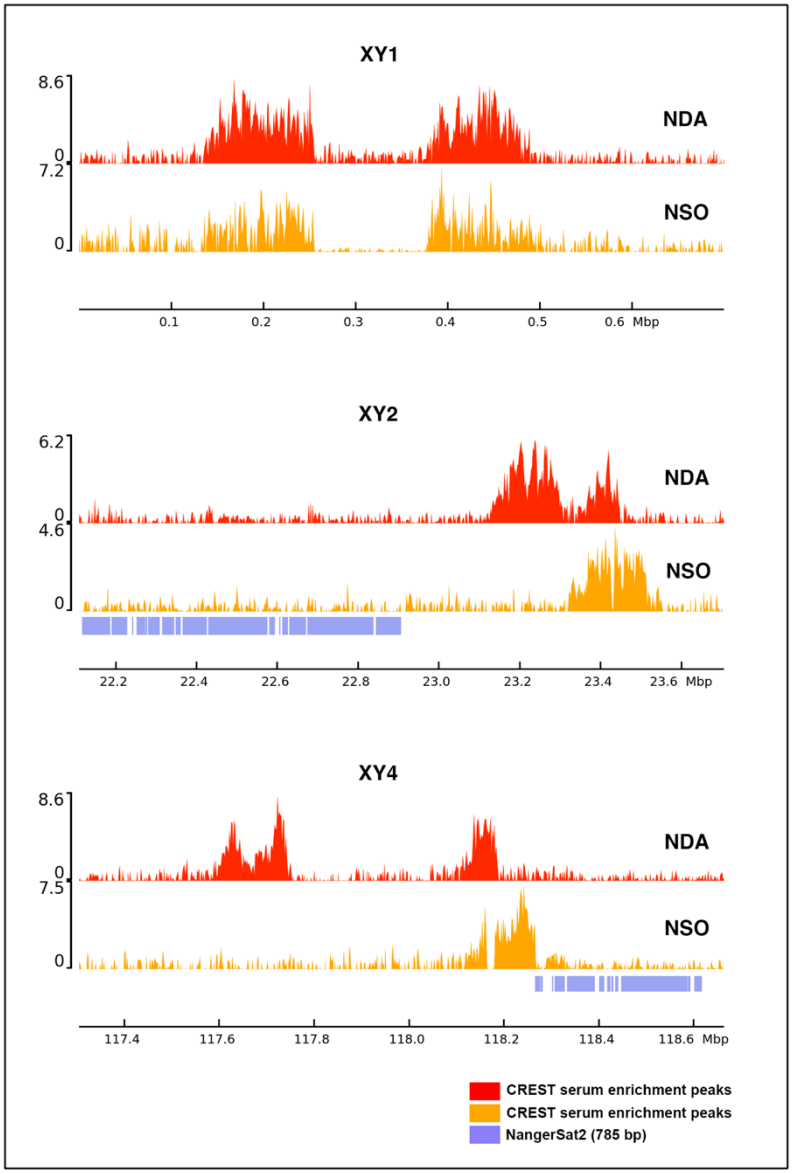

Figure 23. **Satellite-free centromeres at orthologous positions in two gazelle species.** CENP-A ChIP-seq enrichment profiles for chromosomes XY1, XY2, and XY4 in *Nanger dama* (red) and *Nanger soemmerringii* (yellow). The y-axis reports the normalized read counts whereas the x-axis reports the coordinates on gazelle genomic assembly Indigo bars indicate localization of NangerSat2.

4. Identification of chromosomes devoid of satellite DNA in *Nanger* species

The karyotypes of the species belonging to the genus *Nanger* are characterized by significant plasticity and genomic instability, since some autosomal and sexual chromosomes are involved in fusion events that are polymorphic in the population (Vassart *et al.,* 1995; Steiner *et al.,* 2015). In particular, specific chromosomal fusions between sex chromosomes and autosomes (orthologous to chromsomes 5 and 16 of *Bos taurus*) have been identified in both *Nanger dama* and *Nanger soemmerringii* species (**Figure 24A**).

According to the chromosome nomenclature proposed by Vassart *et al.,* (1995), the chromosomes resulting from these fusion events in *Nanger dama* and *Nanger soemmerringii* are named X_1, X_2, Y_1 and Y_2 (**Figure 24A**). This nomenclature describes a system where ancestral sex chromosomes have fused with specific autosomes, which are numbered according to their equivalents in the cattle standard karyotype (ISCNDA 1989).

- **X_1** is a biarmed (submetacentric) chromosome, resulting from a translocation between the ancestral X chromosome and the equivalent of cattle chromosome 5. Specifically, the distal part of the long (q) arm of this chromosome is homoeologous to cattle chromosome 5.
- **Y_1** is also a biarmed chromosome, resulting from a translocation between the ancestral Y chromosome and the equivalent of cattle chromosome 16. Unlike other gazelle species, this translocation in *Nanger* occurs without the addition of extra heterochromatin (Vassart *et al.,* 1995).
- **X_2** is an acrocentric autosome orthologous to cattle chromosome 16. The other chromosome orthologous to cattle 16 is fused to the Y chromosome (forming Y_1).
- **Y_2** is an acrocentric autosome that corresponds to cattle chromosome 5. The homologous chromosome is fused with X chromosome.

The cytogenetic evidence obtained from FISH experiments indicates the presence of satellite-free centromeres in three chromosomes in both species (**Figures 18-19**). Furthermore, using ChIP-seq approach, we confirmed at the molecular level the presence of three satellite-free centromeres, placed in the chromosomal sequences named XY1, XY2, and XY4 of *N. dama* primary assembly provided by the RT2T Consortium (**Figures 21-22**). Nevertheless, the “XY” sequences of this assembly lack exact chromosomal numbers. Specifically, the primary haplotype contains four 'XY' sequences (XY1, XY2, XY3, and XY4), whereas three are present in the secondary haplotype (XY1, XY2, and XY3).

To define the orthologies between *Nanger dama* and *Bos taurus* chromosomes, we performed comparative genomic analyses using Chromeister (Afgan *et al.,* 2016). We compared the two *Nanger*

dama haplotypes with the *Bos taurus* reference genome (BosTau9). *Bos taurus* was chosen as a reference because it is a phylogenetically distant ruminant whose karyotype closely resembles the common ancestor of the Bovidae family. The pairwise genome comparisons are not reported here, since in this work we focused on *N. dama* XY chromosomal sequences.

The orthologies between the chromosomal sequences (XY1, XY2, XY3 and XY4) of both *N. dama* haplotypes (x-axis) and *Bos taurus* chromosomes 5, 16, X, and Y (y-axis) are summarized in the Chromeister plot shown in **Figure 24B**. Each segment in the plot indicates high sequence similarity between the two sequences compared. A segment spanning an entire section of the plot indicates completely collinear chromosomes. Conversely, segments covering only a portion of a section indicate partial sequence identity. Furthermore, a descending line indicates that the two aligned regions share the same orientation, whereas an ascending segment signifies that the sequences are inverted relative to each other.

The sequence XY1 from *N. dama* primary haplotype (83 Mb) shows partial sequence identity with the *B. taurus* X chromosome (q arm). Conversely, XY1 sequence from the secondary haplotype (188 Mb) exhibits sequence identity with the p arm of *B. taurus* X chromosome and with the entire cattle 5 chromosome. Conversely, the XY4 sequence from the primary haplotype (118 Mb) is completely orthologous to *B. taurus* chromosome 5.

The XY2 sequence from the primary haplotype (83 Mb) is orthologous to most of *B. taurus* chromosome 16, while the sequence XY2 from the secondary haplotype (29 Mb) shows sequence identity with a short region of *B. taurus* chromosome 16. The XY3 sequence from the primary haplotype (105 Mb) is almost entirely orthologous to *B. taurus* chromosome 16.

Finally, XY3 sequence from the secondary haplotype (110 Mb) does not exhibit a clear sequence identity with none of *B. taurus* chromosomes.

On the basis of these results, we precisely identified in the *N.dama* assembly the chromosomal sequences contained in the XY1, XY2, XY3 and XY4 that were previously described cytogenetically (Vassart 1995). A detailed description of this new molecular characterization is reported in the Discussion.

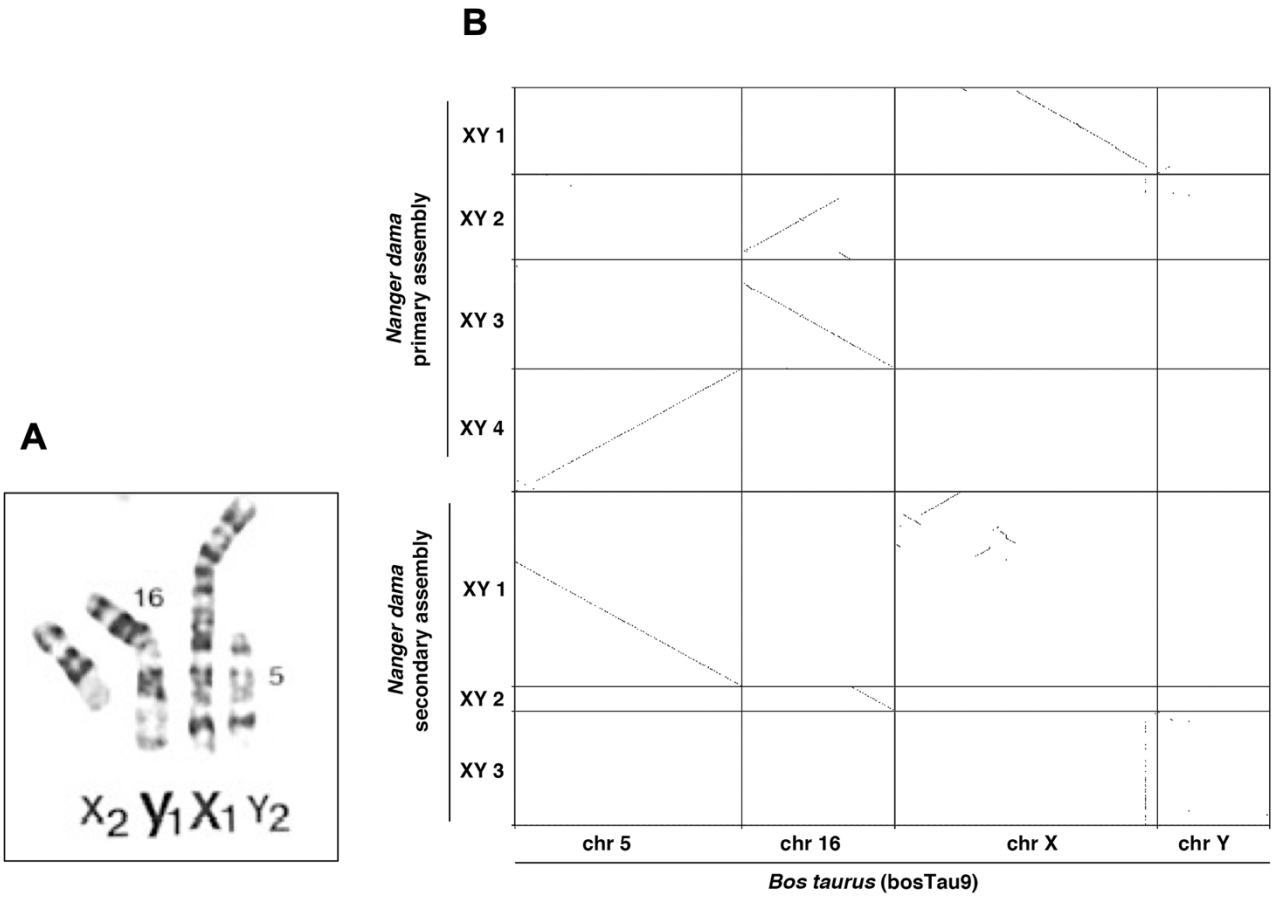

Figure 24. (A) *Nanger dama* chromosomes named X_1, X_2, Y_1 and Y_2, according to the chromosome nomenclature from Vassart et al., 1995. The orthologies with *B. taurus* chromosomes are reported. **(B)** Comparative genomic analyses between the chromosomal sequences XY1, XY2, XY3, and XY4 from both *Nanger dama* haplotypes and *Bos taurus* chromosomes 5, 16, X, and Y. The plot, generated using Chromeister, displays *Bos taurus* chromosomes on the x-axis and *Nanger dama* chromosomal sequences on the y-axis.

5. Cytogenetic and molecular distribution of telomeric repeats in two *Nanger* species

We investigated the distribution of telomeric repeats (TTAGGG)n in both *Nanger dama* and *Nanger soemmerringii* species. To cytogenetically evaluate the chromosomal localization of telomeric repeats in the gazelle species, two FISH experiments were performed on metaphase chromosomes of *Nanger dama* (**Figure 25**) and *Nanger soemmerringii* (**Figure 26**) using a telomeric repeat oligonucleotide probe (Bertoni *et al.,* 1994).

In both species, we observed large blocks of telomeric-like repetitions at the primary constrictions of the majority of chromosomes.

Specifically, in *Nanger dama* (**Figure 25**), intense interstitial telomeric signals were detected at the primary constriction of 27 chromosomes, the majority of which were acrocentric. Furthermore, we identified less intense interstitial signals at the primary constrictions of 7 chromosomes. Based on the analysis of 10 different metaphases, the number of intense centromeric signals ranged from 18 to 28, while the number of less intense signals varied between 3 and 7.

In *Nanger soemmerringii* (**Figure 26**) we observed high-intensity hybridization signals at the primary constrictions of 26 chromosomes, primarily involving the acrocentric chromosomes. Additionally, 6 less intense interstitial signals were observed. The count across 10 different metaphases showed that the number of intense signals at the primary constriction ranged from 23 to 26, while the number of less intense signals varied from 3 to 6.

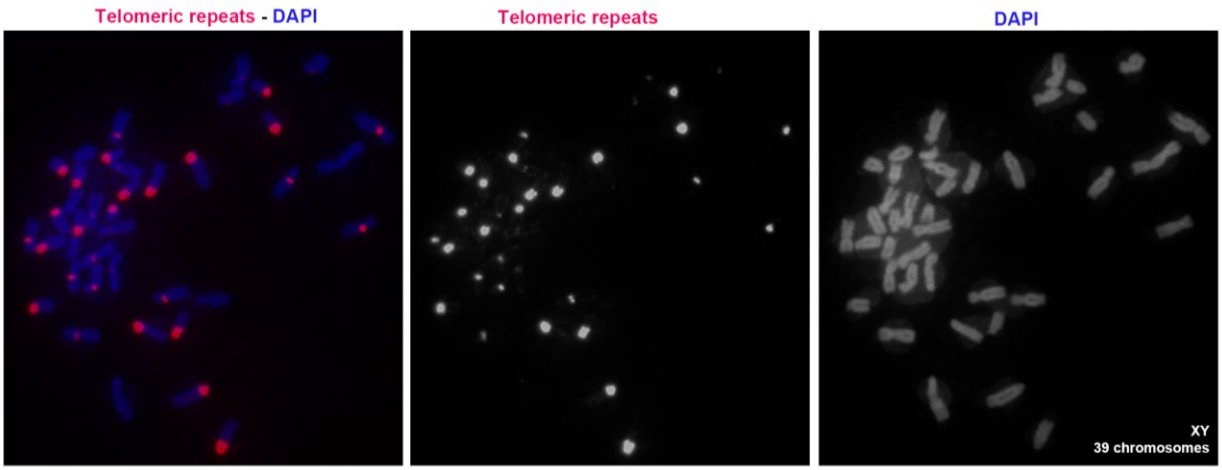

Figure 25. **FISH with a telomeric probe on metaphase chromosomes of *Nanger dama.*** FISH signals (in red) were obtained using a telomeric repeat oligonucleotide probe. The metaphase chromosomes were stained with DAPI (in blue). The high-intensity FISH signals placed in centromeric regions correspond to het-ITSs.

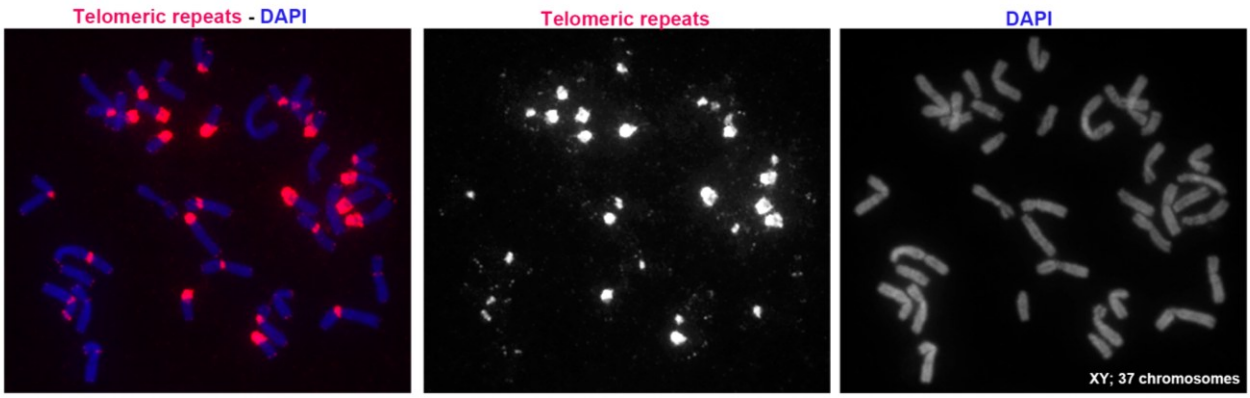

Figure 26. **FISH with a telomeric probe on metaphase chromosomes of *Nanger soemmerringii*.** FISH signals (in red) were obtained using a telomeric repeat oligonucleotide probe. The metaphase chromosomes were stained with DAPI (in blue). The high-intensity FISH signals placed in centromeric regions correspond to het-ITSs.

To confirm the cytogenetic evidence, the genomic distribution of telomeric repeats was investigated in the *Nanger dama* genome assembly using a BLASTN (v2.11.0) search for the $(TTAGGG)_4$ motif. In **Figure 27**, the orthologous chromosomes of *Nanger dama* (NDA) and *Nanger soemmerringii* (NSO) are shown in pairs for direct comparison. The chromosomal orthologies among *Nanger dama*, *Nanger soemmerringii* and *Bos taurus* (BTA) are summarized in **Table 5**.

As shown in **Figure 13**, large blocks of telomeric repeats (pink bars), ranging from 500 kb to 24.8 Mb, were identified in the majority of chromosomes in both gazelle species. In most cases, these repeats are localized in the pericentromeric regions flanking the CENP-A binding domains.

These large blocks correspond to het-ITSs (heterochromatic Interstitial Telomeric Sequences), extended blocks of heterochromatic telomeric-like repeats spanning several hundred kb.

In several chromosomes (e.g., NDA1-NSO2, NDA9-NSO10), the CENP-A peaks are strictly associated with long arrays of NangerSat1 (green), which in turn are flanked by extensive telomeric-like repeats (pink).

Furthermore, several centromeric domains exhibit a complex organization in which NangerSat1 (green), NangerSat2 (indigo), and telomeric repeats (pink) appear interspersed in an alternating pattern. Representative examples of this structural arrangement are observed in NDA4/NSO5 and, to a lesser extent, in NDA6/NSO7 and NDA8/NSO9.

NDA	NSO	BTA
1	2	2+15
2	3	21+8
3	4	22+7
4	5	17+13
5	6	27+12
6	7	28+4
7	8	25+6
8	9	23+10
9	10	24+14
10	11	26+9
11	1q	1
12	12	3
13	13	11
14	14	18
15	15	19
16	16	20
17	1p	29
X2	X2	16
Y1	Y1	16+Y
X1	X1	5+X
Y2	Y2	5

Table 5. **Chromosomal orthology between *Nanger dama* (NDA), *Nanger soemmerringii* (NSO), and *Bos taurus* (BTA).** Orthologies are based on the cattle standard karyotype. The plus sign (+) indicates chromosomal fusions in *Nanger* species involving two *Bos taurus* chromosomes. Sex chromosome nomenclature follows Vassart et al., 1995.

Figure 27. ***Localization of het-ITSs in the pericentromeric regions of Nanger dama and N. soemmerringii chromosomes***. For each centromere, the ChIP-seq enrichment profiles obatined with CREST serum are shown (red for NDA and yellow for NSO). The *y*-axis represents normalized read counts, while the *x*-axis indicates the coordinates on the *N. dama* genomic assembly. The colored bars beneath the profiles denote the distribution of satellite DNA arrays (NangerSat1 in green, NangerSat2 in violet) and the blocks of Interstitial Telomeric Sequences (ITSs, in pink).

DISCUSSION

The centromere is a fundamental nucleoprotein structure of the eukaryotic chromosome, essential for maintaining genomic integrity. Canonical centromeric loci are typically characterized by highly repetitive and rapidly evolving arrays of satellite DNA, which have historically hindered molecular analyses. Nevertheless, during evolution, new and functionally active centromeres ("neocentromeres") can arise at novel chromosomal loci devoid of repetitive DNA sequences, following chromosomal rearrangements. This discovery has been a breakthrough in the understanding of centromeric function. Our laboratory has contributed significantly to this field, by identifying the first satellite-free centromere fixed within a vertebrate species in horse chromosome 11 (Wade et al., 2009). Subsequently, our research group identified several satellite-free neocentromeres across the genus *Equus*, that has proven to be an exceptional model system for studying mammalian chromosomal evolution due to its rapid evolution and karyotype reshuffling, driven by chromosomal rearrangements such as centromere repositioning and Robertsonian fusions. Since these neocentromeres are evolutionarily "young," they have not yet accumulated satellite DNA, making equids an ideal system for investigating how centromeres form and mature over time. Recently, we investigated centromere organization in other species within the order Perissodactyla. Among them, *Tapirus indicus* presents an unconventional centromere landscape with three satellite-free centromeres originated from centromere repositioning events, suggesting that the lack of satellite DNA at centromeres is more widespread than previously thought.

In this work, we investigated whether evolutionary recent satellite-free centromeres are also present in the order Artiodactyla, specifically in two gazelle species within the genus *Nanger* (*N. dama* and *N. soemmerringii*), whose karyotypes are characterized by high genomic instability. By combining cytogenetic evidence (FISH) and molecular data (ChIP-seq), we could identify and characterize canonical satellite-based and satellite-free centromeres that coexist in both species under study. Furthermore, we could assign correct chromosomal numbers to the uncharacterized chromosomal sequences (named XY1, XY2, XY3, XY4) within the draft genome assembly provided by the RT2T Consortium. Finally, we identified heterochromatic Interstitial Telomeric Sequences (het-ITSs) in *Nanger dama* pericentromeric domains at the cytogenetic (FISH) and molecular (BLAST search) level.

Together, these findings provide a first description of centromere organization and evolution in two gazelle species, and notably, extend beyond the order Perissodactyla, which has been the primary focus of our research group so far.

1. Cytogenetic and molecular characterization of satellite-based and satellite-free centromeres in *Nanger dama* and *Nanger soemmerringii*

We focused on two gazelle species belonging to the genus *Nanger* (subfamily Antilopinae), *N. dama* and *N. soemmerringii.* These species exhibit highly variable diploid chromosome numbers (2n = 30–58). Their karyotypes consist predominantly of metacentric chromosomes originated from centric fusions of the 58 acrocentric chromosomes found in the putative ancestral karyotype (Vassart *et al.,* 1995; Cernohorska *et al.,* 2012). Moreover, the karyotypes of these species are extremely plastic and characterized by high genomic instability. In particular, autosomal and sexual chromosomes can be involved in fusion events that are polymorphic in the population (Steiner et al. 2015).

In this study, we analysed a male individual of the species *Nanger dama* (2n = 39) and a male individual of the specis *Nanger soemmerringii* (2n = 37), both characterized by chromosomal fusions involving autosomal and sex chromosomes. Given the high degree of genomic instability and plasticity in these species, we investigated the centromere and satellite DNA organization in both *Nanger* species. Through cytogenetic and molecular approaches, we demonstrated that both possess satellite-free centromeres.

By utilizing genomic reads derived from two distinct ChIP-seq experiments conducted with CREST serum on primary fibroblasts from *N. dama* and *N. soemmerringii*, we identified two distinct satellite DNA families using TAREAN, which we named NangerSat1 and NangerSat2. NangerSat1 (701 bp) is the major centromeric satellite in both species and exhibits high sequence identity with the Artiodactyl satellites OSSAT1 (*Ovis aries*) and BTSAT6 (*Bos taurus*). In contrast, NangerSat2 (785 bp) is found exclusively in *Nanger dama* genomic reads, is not centromeric, and shows significant conservation with OOSAT2 and BTSAT4 from sheep and cattle, respectively.

Since NangerSat1 is shared between the two species under study and is centromeric in both cases, we performed two FISH experiments using as a probe a genomic fragment containing three NangerSat1 repeats on metaphase chromosomes from both gazelles. Both experiments showed that hybridization signals—corresponding to satellite DNA loci—are not present at all primary constrictions. Both species exhibit three chromosomes lacking a detectable signal, specifically two submetacentric and one acrocentric.

However, the resolution of FISH experiments does not allow to exclude the presence of short arrays of satellite DNA at the chromosomal loci that lack clear hybridization signal. Conversely, the presence of a hybridization signal at the primary constriction does not exclude the existence of a satellite-free centromere close to the satellite array: in Burchell's and Grevy's zebras, it has been observed that 11 chromosomes (across both species) show satellite signals at the primary constriction,

while the actual centromeric function has shifted away from the satellite arrays remaining after a centric fusion event (Piras et al., 2010; Cappelletti *et al.,* 2022).

Consequently, cytogenetic evidence must be integrated with molecular data to confirm the presence of satellite-free and satellite-based centromeric loci at the DNA sequence level.

We performed two distinct ChIP-seq experiments on the chromatin extracted from the primary fibroblasts of both species using a CREST serum containing antibodies against centromeric proteins, including the centromere-specific histone H3 variant CENP-A. The ChIP-seq approach allowed us to accurately map centromeric domains of both species and examine their sequence composition. This analysis was made possible by the availability of a preliminary diploid assembly of the *N. dama* genome, provided by the Ruminant Telomere-to-Telomere (RT2T) Consortium. The more complete and higher quality "primary" haplotype of this assembly served as a reference to visualize the localization of satellite DNA arrays and CREST enrichment peaks. Although the assembly consists of chromosome-level scaffolds, it includes sequences designated as XY1, XY2, XY3 and XY4, which, at this stage of the assembly, lack definitive chromosomal numbering but represent distinct genomic loci.

In *N. dama,* we identified 13 canonical satellite-based centromeres and 3 satellite-free centromeres, the latter located on the chromosomal sequences designated as XY1, XY2, and XY4. CREST enrichment peaks were not detected for 5 out of 21 chromosomes (11, 12, 15, 17 and XY3).

In *N. soemmerringii,* we identified 14 canonical satellite-based centromeres and 3 satellite-free centromeres (XY1, XY2, and XY4). CREST enrichment peaks were not detected for 4 out of 21 chromosomes (12, 15, 17 and XY3).

Satellite-based centromeres are embedded within large arrays of NangerSat1, while the NangerSat2 satellite is located in the pericentromeric region, surrounding both satellite-based and satellite-free centromeric domains.

The satellite-based enrichment peaks show irregular shapes, and the underlying satellite DNA arrays appear fragmented by other repetitive sequences not previously detected by TAREAN analysis. This peculiar centromeric organization could be due to errors in the genome assembly, but this is currently under investigation in our laboratory.

In contrast, CREST enrichment peaks at satellite-free centromeres exhibit regular, Gaussian-like shapes, indicating a more accurate underlying sequence assembly due to the absence of repetitive satellite DNA.

The identification of satellite-free centromeres on three chromosomes in the genus *Nanger* is a significant finding, as it represents the first documented example within the order Artiodactyla. Notably, our comparative analysis revealed a striking collinearity between *N.*

dama and *N. soemmerringii*, as these satellite-free domains are located in orthologous genomic positions across all three chromosomes (XY1, XY2, and XY4). This conservation suggests that these genomic regions have features that make them more prone to CENP-A binding and neocentromere formation. In XY1 of both species and XY4 of *Nanger dama*, we detected two distinct enrichment peaks, which may correspond to separate epialleles (Purgato *et al.*, 2015). Differences between homologs in the position of CENP-A binding domains—a phenomenon referred to as epialleles—were originally documented in the horse by our group (Purgato *et al.*, 2015). While some minor differences in peak morphology and number were observed—such as the presence of two distinct peaks in *N. dama* XY4 compared to a single peak in *N. soemmerringii*—the overall centromeric localization remains strictly conserved.

The presence of these epialleles further confirms that centromeric function is determined epigenetically by CENP-A, rather than by the underlying DNA sequence itself.

2. Identification of the chromosomal sequences XY1, XY2, XY3, and XY4 according to Vassart nomenclature

The species belonging to genus *Nanger* are characterized by remarkable genomic plasticity. As previously reported, these species are characterized by complex sex-autosome and autosome-autosome translocations that are polymorphic within populations (Vassart et al., 1995; Steiner et al., 2015).

The chromosome nomenclature proposed by Vassart et al. (1995) for *Nanger dama* and *N. soemmerringii* karyotypes describes a chromosome system comprising the chromosomes named X_1, X_2, Y_1 andY_2. The submetacentric X_1 is the result of the translocation of the ancestral X chromosome to the equivalent of cattle chromosome 5. This chromosomal rearrangement is common among Gazelles and is shared with other Antilopinae (Vassart et al. 1995). The submetacentric chromosome named Y_1 originated from the translocation of the ancestral Y to the equivalent of cattle chromosome 16. Finally, chromosomes X_2 and Y_2 (submetacentric and acrocentric, respectively) are both autosomal and are orthologous to cattle chromosomes 16 and 5, respectively.

A significant limitation in the study of centromere architecture in the gazelle species under study is the lack of specific chromosomal assignments for the chromosomes carrying satellite-free centromeres in the genome assembly provided by the RT2T Consortium, where sex-related sequences are generically labelled as "XY" chromosomes. Specifically, we detected satellite-free centromeres in chromosomes XY1, XY2 and XY4 in both gazelle species.

Performing genomic comparative analyses between RT2T genome assembly and *Bos taurus* reference genome and integrating this with Vassart standard nomenclature, we have been able to assign chromosome numbers for the XY1, XY2, XY3 and XY4 sequences of both haplotypes.

We propose the following chromosome assignments and corrections to the genome assembly:

- Since Chromeister comparative analysis revealed that the XY1 (primary haplotype) shows partial identity with *B. taurus* X chromosome, while the XY1 (secondary haplotype) exhibits identity with both *B. taurus* X and entire BTA5, these sequences are likely part of a single, large submetacentric chromosome. According to Vassart nomenclature, this chromosome is consistent with the X_1. Notably, our identification of a satellite-free centromeres within XY1 sequence suggests that X_1 may harbor one of the three non-canonical centromeric domains.
- The XY4 sequence (primary) is entirely collinear with *Bos taurus* chromosome 5. Based on this, it can be putatively assigned as the acrocentric autosome Y_2, the unfused autosomal partner of the BTA5-homologous portion of X_1. The presence of a satellite-free centromere domain within XY4 sequence, suggests that Y_2 also lacks canonical satellite DNA at its primary constriction.

- The XY3 sequence (primary) shows collinearity with Bos taurus chromosome 16, and it likely corresponds to the autosome X_2, equivalent to BTA16.
- The XY2 (primary) and XY2 (secondary), together with XY3 (secondary) likely correspond to the chromosome Y_1, which originated from a Y-BTA16 translocation. This hypothesis is strongly supported by our preliminary data showing that XY3 (secondary) harbors the *SRY* (Sex-determining Region Y) gene—a master transcription factor that acts as the primary molecular trigger for male sex determination in mammals (Sinclair et al., 1990; Graves, 2006). Furthermore, our ChIP-seq and FISH analyses identified the third satellite-free centromere within the XY2 (primary) sequence. Since XY2 (primary) is orthologous to *Bos taurus* chromosome 16, this finding suggests that the Y_1 chromosome carries a non-canonical, satellite-free centromere at the junction or within its autosomal-derived component.

The lack of sequence identity between XY3 sequence (secondary) and *Bos taurus* Y chromosome can be explained by the presence of species-specific satellite DNA arrays in both Y chromosomes. In mammals, the Y chromosome is known to accumulate vast amounts of repetitive elements and satellite DNA, which undergo rapid evolutionary turnover and structural reorganization (Charlesworth et al., 2005; Bachtrog, 2013). Such fast-evolving sequences often create significant barriers to cross-species genomic alignments, particularly in Bovidae, where sex-specific satellite families can differ substantially even between closely related genera (Jobse et al., 1995; Paria et al., 2011).

In conclusion, we proposed that the chromosomes X_1, Y_1 and Y_2 carry the three satellite-free centromeres identified by FISH and ChIP-seq approaches.

However, the identification of gazelle chromosome carrying satellite-free centromeres remains provisional. The definitive assignment of "XY" sequences to the X_1, X_2, Y_1 or Y_2 complex can only be confirmed through a complete reconstruction of the karyotype of the *Nanger dama* individual used for sequencing and genome assembly.

3. Evolutionary significance of pericentromeric Interstitial Telomeric Sequences (ITSs)

The karyotype evolution of the genus *Nanger* is marked by a high frequency of centric fusions (Vassart et al., 1995), a process that typically leaves detectable molecular traces at the rearrangement sites. Our investigation into the distribution of telomeric repeats (TTAGGG)n in both *N. dama* genome assembly provides compelling evidence of these ancestral events. Cytogenetic analysis via FISH, performed using a specific telomeric repeat oligonucleotide probe (TTAGGG)n (Bertoni et al., 1994), revealed high-intensity hybridization signals at the primary constrictions of the majority of chromosomes in both species (25 chromosomes in *N. dama* and 28 in *N. soemmerringii*). To confirm this evidence at the DNA sequence level, we conducted a BLASTN search for the $(TTAGGG)_4$ motif within the *Nanger dama* assembly. This genomic approach identified extensive telomeric arrays, ranging from 500 kb up to 24.8 Mb, primarily placed in pericentromeric regions.

These large blocks may correspond to het-ITSs (heterochromatic Interstitial Telomeric Sequences), extended blocks of heterochromatic telomeric-like repeats spanning several hundred kb. These sequences are mainly localized in centromeric or pericentromeric chromosome regions in various vertebrate species (Meyne et al., 1990; Ruiz-Herrera et al., 2008; Bolzán 2017). It has been proposed that het-ITSs are likely remnants of ancestral chromosomal rearrangements produced during the evolution of karyotypes in different taxa. Specifically, they may correspond to the telomeres retained after a fusion event between two chromosomal ends (Ruiz-Herrera et al., 2008).

In the genus *Nanger*, the massive presence of these telomeric-like sequences—often interspersed with satellite DNA families NangerSat1 and NangerSat2—suggests they are not merely remnants but integral components of the centromeric heterochromatic environment. This complex sequence organization, particularly evident in orthologous pairs such as NDA1-NSO2, NDA4-NSO5, or NDA9-NSO10, suggests that telomeric repeats may play a role in stabilizing new centromeric regions following Robertsonian translocations. Consequently, the abundance of het-ITSs in *Nanger dama* serves as a molecular signature of the extensive reorganization that shaped the *Nanger* genome starting from the ancestral Bovidae karyotype.

CONCLUSIONS

This study provides a first characterization of the centromeric organization in two gazelle species belonging to the genus *Nanger* (*N. dama* and *N. soemmerringii*) offering new insights into the mechanisms of genomic evolution within the order Artiodactyla. The integration of cytogenetic and molecular approaches allowed us to obtain the following results:

1. We identified the major centromeric satellite DNA family (NangerSat1, 701 bp long), shared by both *Nanger* species. Extended NangerSat1 arrays were identified in the majority of centromeric domains in the two gazelle species.
2. Both species have three functionally active centromeres entirely devoid of satellite DNA and localized at orthologous positions.
3. Preforming genomic comparative analyses between the draft genome assembly of *Nanger dama* provided by the RT2T and *Bos taurus* reference genome, we propose that the three satellite-free centromeres are located on chromosomes X_1, Y_1, both originated from chromosomal fusion events, and Y_2.
4. Extensive heterochromatic Interstitial Telomeric Sequences (het-ITSs) are placed in the pericentromeric regions of the majority of *N. dama* chromosomes. These sequences appear to be integral components of the centromeric environment and reveal the massive reorganization that shaped *Nanger* genomes from the ancestral Bovidae karyotype.

In conclusion, the high rate of genomic plasticity led to the formation of satellite-free unconventional centromere architectures in the *Nanger* species under study. The lack of satellite DNA at centromeres is not a feature restricted to Perissodactyla order, but more widespread than previously thought among mammalian species.

BIBLIOGRAPHY

Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Res*. 2016 Jul 8;44(W1): W3-W10.

Allshire RC, Karpen GH. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat Rev Genet*. 2008 Dec;9(12):923-37.

Amor DJ, Bentley K, Ryan J, Perry J, Wong L, Slater H, Choo KH. Human centromere repositioning "in progress". *Proc Natl Acad Sci U S A*. 2004 Apr 27;101(17):6542-7.

Amor DJ, Choo KH. Neocentromeres: role in human disease, evolution, and centromere study. *Am J Hum Genet*. 2002 Oct;71(4):695-714.

Bachtrog D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet*. 2013 Feb;14(2):113-24.

Bergmann JH, Rodríguez MG, Martins NM, Kimura H, Kelly DA, Masumoto H, Larionov V, Jansen LE, Earnshaw WC. Epigenetic engineering shows H3K4me2 is required for HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J.* 2011 Jan 19;30(2):328-40.

Bertoni L, Attolini C, Tessera L, Mucciolo E, Giulotto E. Telomeric and nontelomeric (TTAGGG)n sequences in gene amplification and chromosome stability. *Genomics*. 1994 Nov 1;24(1):53-62.

Black BE, Brock MA, Bédard S, Woods VL Jr, Cleveland DW. An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes. *Proc Natl Acad Sci USA*. 2007 Mar 20;104(12):5008-13.

Blower MD, Sullivan BA, Karpen GH. Conserved organization of centromeric chromatin in flies and humans. *Dev Cell*. 2002 Mar;2(3):319-30.

Bobeica C, Niculet E, Craescu M, Parapiru EL, Musat CL, Dinu C, Chiscop I, Nechita L, Debita M, Stefanescu V, Stefanopol IA, Nechifor A, Pelin AM, Balan G, Chirobocea S, Vasile CI, Tatu

AL. CREST Syndrome in Systemic Sclerosis Patients - Is Dystrophic Calcinosis a Key Element to a Positive Diagnosis? *J Inflamm Res*. 2022 Jun 9; 15:3387-3394.

Bobkov GOM, Gilbert N, Heun P. Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. *J Cell Biol*. 2018 Jun 4;217(6):1957-1972.

Bolzán AD. Interstitial telomeric sequences in vertebrate chromosomes: Origin, function, instability and evolution. *Mutat Res Rev Mutat Res*. 2017 Jul; 773:51-65.

Capozzi O, Purgato S, Verdun di Cantogno L, Grosso E, Ciccone R, Zuffardi O, Della Valle G, Rocchi M. Evolutionary and clinical neocentromeres: two faces of the same coin? *Chromosoma*. 2008 Aug;117(4):339-44.

Cappelletti E, Piras FM, Badiale C, Bambi M, Santagostino M, Vara C, Masterson TA, Sullivan KF, Nergadze SG, Ruiz-Herrera A, Giulotto E. CENP-A binding domains and recombination patterns in horse spermatocytes. *Sci Rep*. 2019 Nov 1;9(1):15800.

Cappelletti E, Piras FM, Biundo M, Raimondi E, Nergadze SG, Giulotto E. CENP-A/CENP-B uncoupling in the evolutionary reshuffling of centromeres in equids. *Genome Biol*. 2025 Feb 6;26(1):23.

Cappelletti E, Piras FM, Sola L, Santagostino M, Abdelgadir WA, Raimondi E, Lescai F, Nergadze SG, Giulotto E. Robertsonian Fusion and Centromere Repositioning Contributed to the Formation of Satellite-free Centromeres During the Evolution of Zebras. *Mol Biol Evol*. 2022 Aug 3;39(8):msac162.

Carbone L, Nergadze SG, Magnani E, Misceo D, Francesca Cardone M, Roberto R, Bertoni L, Attolini C, Francesca Piras M, de Jong P, Raudsepp T, Chowdhary BP, Guérin G, Archidiacono N, Rocchi M, Giulotto E. Evolutionary movement of centromeres in horse, donkey, and zebra. *Genomics*. 2006 Jun;87(6):777-82.

Cardone MF, Alonso A, Pazienza M, Ventura M, Montemurro G, Carbone L, de Jong PJ, Stanyon R, D'Addabbo P, Archidiacono N, She X, Eichler EE, Warburton PE, Rocchi M. Independent centromere formation in a capricious, gene-free domain of chromosome 13q21 in Old World monkeys and pigs. *Genome Biol*. 2006;7(10):R91.

Carroll CW, Milks KJ, Straight AF. Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J. Cell Biol*. 2010 Jun 28;189(7):1143-55.

Cernohorska H, Kubickova S, Vahala J, Rubes J. Molecular insights into X; BTA5 chromosome rearrangements in the tribe Antilopini (Bovidae). *Cytogenet Genome Res*. 2012;136(3):188 98.

Chan FL, Wong LH. Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Res*. 2012 Dec;40(22):11178-88.

Chardon F, Japaridze A, Witt H, Velikovsky L, Chakraborty C, Wilhelm T, Dumont M, Yang W, Kikuti C, Gangnard S, Mace AS, Wuite G, Dekker C, Fachinetti D. CENP-B- mediated DNA loops regulate activity and stability of human centromeres. *Mol Cell*. 2022 May 5;82(9):1751-1767.e8.

Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. *Heredity (Edinb)*. 2005 Aug;95(2):118-28.

Choo KH. Centromere DNA dynamics: latent centromeres and neocentromere formation. *Am J Hum Genet*. 1997 Dec;61(6):1225-33.

Choo KH. Centromerization. *Trends Cell Biol*. 2000 May;10(5):182-8.

Clarke L. Centromeres: proteins, protein complexes, and repeated domains at centromeres of simple eukaryotes. *Curr Opin Genet Dev*. 1998 Apr;8(2):212-8.

Cleveland DW, Mao Y, Sullivan KF. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell*. 2003 Feb 21;112(4):407-21.

Corless S, Höcker S, Erhardt S. Centromeric RNA and Its Function at and Beyond Centromeric Chromatin. *J Mol Biol*. 2020 Jul 10;432(15):4257-4269.

de Groot C, Houston J, Davis B, Gerson-Gurwitz A, Monen J, Lara-Gonzalez P, Oegema K, Shiau AK, Desai A. The N-terminal tail of C. elegans CENP-A interacts with KNL-2 and is essential for centromeric chromatin assembly. *Mol Biol Cell.* 2021 Jun 1;32(12):1193-1201.

DeBose-Scarlett EM, Sullivan BA. Genomic and Epigenetic Foundations of Neocentromere Formation. *Annu Rev Genet*. 2021 Nov 23;55:331-348.

Depinet TW, Zackowski JL, Earnshaw WC, Kaffe S, Sekhon GS, Stallard R, Sullivan BA, Vance GH, Van Dyke DL, Willard HF, Zinn AB, Schwartz S. Characterization of neo-centromeres in marker chromosomes lacking detectable alpha-satellite DNA. *Hum Mol Genet*. 1997 Aug;6(8):1195-204.

Dong Q, Li F. Cell cycle control of kinetochore assembly. *Nucleus*. 2022 Dec;13(1):208-220.

Dover G. Molecular drive: a cohesive mode of species evolution. *Nature*. 1982 Sep 9;299(5879):111-7.

Dudka D, Lampson MA. Centromere drive: model systems and experimental progress. *Chromosome Res*. 2022 Sep;30(2-3):187-203.

Earnshaw WC, Migeon BR. Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma*. 1985;92(4):290-6.

Earnshaw WC, Rothfield N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma.* 1985;91(3-4):313-21.

Eissenberg JC, James TC, Foster-Hartnett DM, Hartnett T, Ngan V, Elgin SC. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. *Proc Natl Acad Sci U S A*. 1990 Dec;87(24):9923-7.

Elder JF Jr, Turner BJ. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q Rev Biol*. 1995 Sep;70(3):297-320.

Fachinetti D, Han JS, McMahon MA, Ly P, Abdullah A, Wong AJ, Cleveland DW. DNA Sequence-Specific Binding of CENP-B Enhances the Fidelity of Human Centromere Function. *Dev Cell*. 2015 May 4;33(3):314-27.

Foltz DR, Jansen LE, Bailey AO, Yates JR 3rd, Bassett EA, Wood S, Black BE, Cleveland DW. Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell*. 2009 May 1;137(3):472-84.

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 Dec;12(4):1069-84.

Fukagawa T, Earnshaw WC. The centromere: chromatin foundation for the kinetochore machinery. *Dev Cell*. 2014 Sep 8;30(5):496-508.

Gamba R, Fachinetti D. From evolution to function: Two sides of the same CENP-B coin? *Exp Cell Res*. 2020 May 15;390(2):111959.

Garrido-Ramos MA. Satellite DNA: An Evolving Topic. *Genes (Basel)*. 2017 Sep 18;8(9):230.

Giulotto E, Raimondi E, Sullivan KF. The Unique DNA Sequences Underlying Equine Centromeres. *Prog Mol Subcell Biol*. 2017; 56:337-354.

Goutte-Gattat D, Shuaib M, Ouararhni K, Gautier T, Skoufias DA, Hamiche A, Dimitrov S. Phosphorylation of the CENP-A amino-terminus in mitotic centromeric chromatin is required for kinetochore function. *Proc Natl Acad Sci USA*. 2013 May 21;110(21):8579-84.

Graves JA. Sex chromosome specialization and degeneration in mammals. *Cell*. 2006 Mar 10;124(5):901-14.

Hamilton GE, Davis TN. Biochemical evidence for diverse strategies in the inner kinetochore. *Open Biol*. 2020 Nov;10(11):200284.

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet*. 1997 Apr;15(4):345-55.

Hartley G, O'Neill RJ. Centromere Repeats: Hidden Gems of the Genome. *Genes (Basel)*. 2019 Mar 16;10(3):223.

Hartley GA, Okhovat M, Hoyt SJ, Fuller E, Pauloski N, Alexandre N, Alexandrov I, Drennan R, Dubocanin D, Gilbert DM, Mao Y, McCann C, Neph S, Ryabov F, Sasaki T, Storer JM, Svendsen D, Troy W, Wells J, Core L, Stergachis A, Carbone L, O'Neill RJ. Centromeric transposable elements and epigenetic status drive karyotypic variation in the eastern hoolock gibbon. *Cell Genom*. 2025 Apr 9;5(4):100808.

Hassold T, Hunt P. To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet*. 2001 Apr;2(4):280-91.

Hasson D, Alonso A, Cheung F, Tepperberg JH, Papenhausen PR, Engelen JJ, Warburton PE. Formation of novel CENP-A domains on tandem repetitive DNA and across chromosome breakpoints on human chromosome 8q21 neocentromeres. *Chromosoma*. 2011 Dec;120(6):621-32.

Hedouin S, Logsdon GA, Underwood JG, Biggins S. A transcriptional roadblock protects yeast centromeres. *Nucleic Acids Res*. 2022 Aug 12;50(14):7801-7815.

Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*. 2001 Aug 10;293(5532):1098-102.

Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, Karpen GH. Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell*. 2006 Mar;10(3):303-15.

Hill A, Bloom K. Genetic manipulation of centromere function. *Mol Cell Biol.* 1987 Jul;7(7):2397-405.

Howman EV, Fowler KJ, Newson AJ, Redward S, MacDonald AC, Kalitsis P, Choo KH. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc Natl Acad Sci U S A*. 2000 Feb 1;97(3):1148-53.

Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H. Construction of YAC-based mammalian artificial chromosomes. *Nat Biotechnol*. 1998 May;16(5):431-9.

Jobse C, Buntjer JB, Haagsma N, Breukelman HJ, Beintema JJ, Lenstra JA. Evolution and recombination of bovine DNA repeats. *J Mol Evol*. 1995 Sep;41(3):277-83.

Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, Albrechtsen A, Petersen B, Korneliussen TS, Vilstrup JT, Lear T, Myka JL, Lundquist J, Miller DC, Alfarhan AH, Alquraishi SA, Al-Rasheid KA, Stagegaard J, Strauss G, Bertelsen MF, Sicheritz-Ponten T, Antczak DF, Bailey E, Nielsen R, Willerslev E, Orlando L. Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci U S A*. 2014 Dec 30;111(52):18655-60.

Kalbfleisch TS, McKay SD, Murdoch BM, Adelson DL, Almansa-Villa D, Becker G, Beckett LM, Benítez-Galeano MJ, Biase F, Casey T, Chuong E, Clark E, Clarke S, Cockett N, Couldrey C, Davis BW, Elsik CG, Faraut T, Gao Y, Genet C, Grady P, Green J, Green R, Guan D, Hagen D, Hartley GA, Heaton M, Hoyt SJ, Huang W, Jarvis E, Kalleberg J, Khatib H, Koepfi KP, Koltes J, Koren S, Kuehn C, Leeb T, Leonard A, Liu GE, Low WY, McConnell H, McRae K, Miga K, Mousel M, Neibergs H, Olagunju T, Pennell M, Petry B, Pewsner M, Phillippy AM, Pickett BD, Pineda P, Potapova T, Rachagani S, Rhie A, Rijnkels M, Robic A, Rodriguez Osorio N, Safonova Y, Schettini G, Schnabel RD, Sirpu Natesh N, Stegemiller M, Storer J, Stothard P, Stull C, Tosser-Klopp G, Traglia GM, Tuggle CK, Van Tassell CP, Watson C, Weikard R, Wimmers K, Xie S, Yang L, Smith TPL, O'Neill RJ, Rosen BD. The Ruminant Telomere-to-Telomere (RT2T) Consortium. *Nat Genet*. 2024 Aug;56(8):1566-1573.

Kalitsis P, Choo KH. The evolutionary life cycle of the resilient centromere. *Chromosoma.* 2012 Aug;121(4):327-40.

Kixmoeller K, Allu PK, Black BE. The centromere comes into focus: from CENP-A nucleosomes to kinetochore connections with the spindle. *Open Biol*. 2020 Jun;10(6):200051.

Kursel LE, Malik HS. The cellular mechanisms and consequences of centromere drive. *Curr Opin Cell Biol*. 2018 Jun;52:58-65.

Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA. Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc Natl Acad Sci U S A*. 2006 Mar 14;103(11):4186-91.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012 Mar 4;9(4):357-9.

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.

Lechner J, Ortiz J. The Saccharomyces cerevisiae kinetochore. *FEBS Lett*. 1996 Jun 24;389(1):70-4.

Legrand M, Jaitly P, Feri A, d'Enfert C, Sanyal K. Candida albicans: An Emerging Yeast Model to Study Eukaryotic Genome Plasticity. *Trends Genet*. 2019 Apr;35(4):292-307.

Levan, A, Fredga, K, Sandberg, AA. Nomenclature for centromeric position on chromosomes. *Hereditas*. 1964; 52: 201-220.

Ling YH, Yuen KWY. Point centromere activity requires an optimal level of centromeric noncoding RNA. *Proc Natl Acad Sci U S A*. 2019 Mar 26;116(13):6270-6279.

Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics*. 2021 Apr 20;37(3):422-423.

Maio JJ. DNA strand reassociation and polyribonucleotide binding in the African green monkey, Cercopithecus aethiops. *J Mol Biol*. 1971 Mar 28;56(3):579-95.

Malik HS, Bayes JJ. Genetic conflicts during meiosis and the evolutionary origins of centromere complexity. *Biochem Soc Trans*. 2006 Aug;34(Pt 4):569-73.

Malik HS, Henikoff S. Major evolutionary transitions in centromere complexity. *Cell*. 2009 Sep 18;138(6):1067-82.

Marshall OJ, Chueh AC, Wong LH, Choo KH. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet*. 2008 Feb;82(2):261-82.

Martins NM, Bergmann JH, Shono N, Kimura H, Larionov V, Masumoto H, Earnshaw WC. Epigenetic engineering shows that a human centromere resists silencing mediated by H3K27me3/K9me3. *Mol Biol Cell*. 2016 Jan 1;27(1):177-96.

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 Nov;109(5):1963-73.

McKinley KL, Cheeseman IM. The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol*. 2016 Jan;17(1):16-29.

Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, Garcia JF, DeRisi JL, Smith T, Tobias C, Ross-Ibarra J, Korf I, Chan SW. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol*. 2013 Jan 30;14(1):R10.

Meyne J, Baker RJ, Hobart HH, Hsu TC, Ryder OA, Ward OG, Wiley JE, Wurster-Hill DH, Yates TL, Moyzis RK. Distribution of non-telomeric sites of the (TTAGGG)n telomeric sequence in vertebrate chromosomes. *Chromosoma*. 1990 Apr;99(1):3-10.

Molina O, Vargiu G, Abad MA, Zhiteneva A, Jeyaprakash AA, Masumoto H, Kouprina N, Larionov V, Earnshaw WC. Epigenetic engineering reveals a balance between histone modifications and transcription in kinetochore maintenance. *Nat Commun*. 2016 Nov 14;7: 13334.

Montefalcone G, Tempesta S, Rocchi M, Archidiacono N. Centromere repositioning. *Genome Res*. 1999 Dec;9(12):1184-8.

Moreno-Moreno O, Torras-Llort M, Azorín F. Proteolysis restricts localization of CID, the centromere-specific histone H3 variant of Drosophila, to centromeres. *Nucleic Acids Res*. 2006;34(21):6247-55.

Nagaki K, Kashihara K, Murata M. Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant Luzula nivea. *Plant Cell*. 2005 Jul;17(7):1886-93.

Nakano M, Cardinale S, Noskov VN, Gassmann R, Vagnarelli P, Kandels-Lewis S, Larionov V, Earnshaw WC, Masumoto H. Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev Cell*. 2008 Apr;14(4):507-22.

Naughton C, Gilbert N. Centromere chromatin structure - Lessons from neocentromeres. *Exp Cell Res*. 2020 Apr 15;389(2):111899.

Nergadze SG, Piras FM, Gamba R, Corbo M, Cerutti F, McCarter JGW, Cappelletti E, Gozzo F, Harman RM, Antczak DF, Miller D, Scharfe M, Pavesi G, Raimondi E, Sullivan KF, Giulotto E. Birth, evolution, and transmission of satellite-free mammalian centromeric domains. *Genome Res*. 2018 Jun;28(6):789-799.

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 Jul 7;45(12):e111.

Ohkuni K, Kitagawa K. Role of transcription at centromeres in budding yeast. *Transcription*. 2012 Jul-Aug;3(4):193-7.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson PL, Fumagalli M, Vilstrup JT, Raghavan M, Korneliussen T, Malaspinas AS, Vogt J, Szklarczyk D, Kelstrup CD, Vinther J, Dolocan A, Stenderup J, Velazquez AM, Cahill J, Rasmussen M, Wang X, Min J, Zazula GD, Seguin-Orlando A, Mortensen C, Magnussen K, Thompson JF, Weinstock J, Gregersen K, Røed KH, Eisenmann V, Rubin CJ, Miller DC, Antczak DF, Bertelsen MF, Brunak S, Al-Rasheid KA, Ryder O, Andersson L, Mundy J, Krogh A, Gilbert MT, Kjær K, Sicheritz-Ponten T, Jensen LJ, Olsen JV, Hofreiter M, Nielsen R, Shapiro B, Wang J, Willerslev E. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. 2013 Jul 4;499(7456):74-8.

Palmer DK, O'Day K, Wener MH, Andrews BS, Margolis RL. A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones. *J. Cell Biol*. 1987;104, 805–815.

Paria N, Raudsepp T, Pearks Wilkerson AJ, O'Brien PC, Ferguson-Smith MA, Love CC, Arnold C, Rakestraw P, Murphy WJ, Chowdhary BP. A gene catalogue of the euchromatic male-

specific region of the horse Y chromosome: comparison with human and other mammals. *PLoS One*. 2011;6(7):e21374.

Peng S, Petersen JL, Bellone RR, Kalbfleisch T, Kingsley NB, Barber AM, Cappelletti E, Giulotto E, Finno CJ. Decoding the Equine Genome: Lessons from ENCODE. *Genes (Basel)*. 2021 Oct 27;12(11):1707.

Perpelescu M, Hori T, Toyoda A, Misu S, Monma N, Ikeo K, Obuse C, Fujiyama A, Fukagawa T. HJURP is involved in the expansion of centromeric chromatin. *Mol Biol Cell*. 2015 Aug 1;26(15):2742-54.

Piras FM, Cappelletti E, Abdelgadir WA, Salamon G, Vignati S, Santagostino M, Sola L, Nergadze SG, Giulotto E. A Satellite-Free Centromere in Equus przewalskii Chromosome 10. *Int J Mol Sci*. 2023 Feb 18;24(4):4134.

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 Apr 10;23(8):4183.

Piras FM, Nergadze SG, Magnani E, Bertoni L, Attolini C, Khoriauli L, Raimondi E, Giulotto E. Uncoupling of satellite DNA and centromeric function in the genus Equus. *PLoS Genet*. 2010 Feb 12;6(2): e1000845.

Plohl M, Meštrović N, Mravinac B. Centromere identity from the DNA point of view. *Chromosoma*. 2014 Aug;123(4):313-25.

Przewloka MR, Glover DM. The kinetochore and the centromere: a working long distance relationship. *Annu Rev Genet*. 2009; 43:439-65.

Purgato S, Belloni E, Piras FM, Zoli M, Badiale C, Cerutti F, Mazzagatti A, Perini G, Della Valle G, Nergadze SG, Sullivan KF, Raimondi E, Rocchi M, Giulotto E. Centromere sliding on a mammalian chromosome. *Chromosoma*. 2015 Jun;124(2):277-87.

Quénet D, Dalal Y. A long non-coding RNA is required for targeting centromeric protein A to the human centromere. *Elife*. 2014 Aug 12;3:e03254.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W160-5.

Rieder CL. Ribonucleoprotein staining of centrioles and kinetochores in newt lung cell spindles. *J Cell Biol*. 1979 Jan;80(1):1-9.

Rocchi M, Archidiacono N, Schempp W, Capozzi O, Stanyon R. Centromere repositioning in mammals. *Heredity (Edinb).* 2012 Jan;108(1):59-67.

Rošić S, Erhardt S. No longer a nuisance: long non-coding RNAs join CENP-A in epigenetic centromere regulation. *Cell Mol Life Sci*. 2016 Apr;73(7):1387-98.

Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E. Telomeric repeats far from the ends: mechanisms of origin and role in evolution. *Cytogenet Genome Res*. 2008;122(3-4):219-28.

Sadeghi L, Siggens L, Svensson JP, Ekwall K. Centromeric histone H2B monoubiquitination promotes noncoding transcription and chromatin integrity. *Nat Struct Mol Biol*. 2014 Mar;21(3):236-43.

Saffery R, Sumer H, Hassan S, Wong LH, Craig JM, Todokoro K, Anderson M, Stafford A, Choo KH. Transcription within a functional human centromere. *Mol Cell.* 2003 Aug;12(2):509-16.

Salser W, Bowen S, Browne D, el-Adli F, Fedoroff N, Fry K, Heindell H, Paddock G, Poon R, Wallace B, Whitcome P. Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Fed Proc*. 1976 Jan;35(1):23-35.

Sekulic N, Bassett EA, Rogers DJ, Black BE. The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. *Nature*. 2010 Sep 16;467(7313):347-51.

Senaratne AP, Cortes-Silva N, Drinnenberg IA. Evolution of holocentric chromosomes: Drivers, diversity, and deterrents. *Semin Cell Dev Biol*. 2022 Jul;127:90-99.

Senaratne AP, Muller H, Fryer KA, Kawamoto M, Katsuma S, Drinnenberg IA. Formation of the CenH3-Deficient Holocentromere in Lepidoptera Avoids Active Chromatin. *Curr Biol*. 2021 Jan 11;31(1):173-181.e7.

Senn H, Banfield L, Wacher T, Newby J, Rabeil T, Kaden J, Kitchener AC, Abaigar T, Silva TL, Maunder M, Ogden R. Splitting or lumping? A conservation dilemma exemplified by the critically endangered dama gazelle (*Nanger dama*). *PLoS One*. 2014 Jun 23;9(6):e98693.

Shang WH, Hori T, Toyoda A, Kato J, Popendorf K, Sakakibara Y, Fujiyama A, Fukagawa T. Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res*. 2010 Sep;20(9):1219-28.

Shepelev VA, Uralsky LI, Alexandrov AA, Yurov YB, Rogaev EI, Alexandrov IA. Annotation of suprachromosomal families reveals uncommon types of alpha satellite organization in pericentromeric regions of hg38 human genome assembly. *Genom Data*. 2015 Sep 1;5:139-146.

Shrestha RL, Ahn GS, Staples MI, Sathyan KM, Karpova TS, Foltz DR, Basrai MA. Mislocalization of centromeric histone H3 variant CENP-A contributes to chromosomal instability (CIN) in human cells. *Oncotarget*. 2017 Jul 18;8(29):46781-46800.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature.* 1990 Jul 19;346(6281):240-4.

Smith MM. Centromeres and variant histones: what, where, when and why? *Curr Opin Cell Biol.* 2002 Jun;14(3):279-85.

Steiner CC, Charter SJ, Goddard N, Davis H, Brandt M, Houck ML, Ryder OA. Chromosomal variation and perinatal mortality in San Diego Zoo Soemmerring's gazelles. *Zoo Biol*. 2015 Jul-Aug;34(4):374-84.

Sullivan B.A., G.H. Karpen. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin, *Nat. Struct. Mol. Biol*. 11 (2004) 1076–1083

Sullivan BA, Willard HF. Stable dicentric X chromosomes with two functional centromeres. *Nat Genet*. 1998 Nov;20(3):227-8.

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.* 1 November 1994; 127 (3): 581–592.

Talbert P, Henikoff S. Centromere drive: chromatin conflict in meiosis. *Curr Opin Genet Dev*. 2022 Dec; 77:102005.

Thompson SL, Bakhoum SF, Compton DA. Mechanisms of chromosomal instability. *Curr Biol.* 2010 Mar 23;20(6):R285-95.

Tolomeo D, Capozzi O, Stanyon RR, Archidiacono N, D'Addabbo P, Catacchio CR, Purgato S, Perini G, Schempp W, Huddleston J, Malig M, Eichler EE, Rocchi M. Epigenetic origin of evolutionary novel centromeres. *Sci Rep*. 2017 Feb 3;7:41980.

Trifonov VA, Musilova P, Kulemsina AI. Chromosome evolution in Perissodactyla. *Cytogenet Genome Res*. 2012;137(2-4):208-17.

Trifonov VA, Stanyon R, Nesterenko AI, Fu B, Perelman PL, O'Brien PC, Stone G, Rubtsova NV, Houck ML, Robinson TJ, Ferguson-Smith MA, Dobigny G, Graphodatsky AS, Yang F. Multidirectional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla. *Chromosome Res.* 2008;16(1):89-107.

Vassart M, Greth A, Durand V, Cribiu EP. Chromosomal polymorphism in sand gazelles (Gazella subgutturosa marica). *J Hered*. 1993 Nov-Dec;84(6):478-81.

Vassart M, Séguéla A, Hayes H. Chromosomal evolution in gazelles. *J Hered*. 1995 May-Jun;86(3):216-27.

Voullaire L, Saffery R, Davies J, Earle E, Kalitsis P, Slater H, Irvine DV, Choo KH. Trisomy 20p resulting from inverted duplication and neocentromere formation. *Am J Med Genet*. 1999 Aug 6;85(4):403-8.

Voullaire LE, Slater HR, Petrovic V, Choo KH. A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? Am *J Hum Genet*. 1993 Jun;52(6):1153-63.

Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blöcker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MC, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, Della Valle G, Fryc S, Guérin G, Hasegawa T, Hill EW, Jurka J, Kiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Røed KH, Ryder OA, Searle S, Skow L, Swinburne JE, Syvänen AC, Tozaki T, Valberg SJ, Vaudin M, White JR, Zody MC; Broad Institute Genome Sequencing Platform; Broad Institute Whole Genome Assembly Team; Lander ES, Lindblad-Toh K. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science*. 2009 Nov 6;326(5954):865-7.

Warburton PE, Cooke CA, Bourassa S, Vafa O, Sullivan BA, Stetten G, Gimelli G, Warburton D, Tyler-Smith C, Sullivan KF, Poirier GG, Earnshaw WC. Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr Biol*. 1997 Nov 1;7(11):901-4.

Weaver BA, Cleveland DW. Aneuploidy: instigator and inhibitor of tumorigenesis. *Cancer Res*. 2007;67:10103-10105.

Westermann S, Schleiffer A. Family matters: structural and functional conservation of centromere-associated proteins from yeast to humans. *Trends Cell Biol*. 2013 Jun;23(6):260-9.

Willard HF. Evolution of alpha satellite. *Curr Opin Genet Dev*. 1991 Dec;1(4):509-14.

Yunis JJ, Yasmineh WG. Heterochromatin, satellite DNA, and cell function. Structural DNA of eucaryotes may support and protect genes and aid in speciation. *Science*. 1971 Dec 17;174(4015):1200-9.

Zeng W, Ball AR Jr, Yokomori K. HP1: heterochromatin binding proteins working the genome. *Epigenetics*. 2010 May 16;5(4):287-92.

Zhou B-R, Yadav KNS, Borgnia M, Hong J, Cao B, Olins AL, Olins DE, Bai Y, Zhang P. Atomic resolution cryo-EM structure of a native-like CENP-A nucleosome aided by an antibody fragment. *Nat. Commun*. 2019. 10, 2301

Zhou K, Gebala M, Woods D, Sundararajan K, Edwards G, Krzizike D, Wereszczynski J, Straight AF, Luger K. CENP-N promotes the compaction of centromeric chromatin. *Nat Struct Mol Biol*. 2022 Apr;29(4):403-413.

Zhu J, Guo Q, Choi M, Liang Z, Yuen KWY. Centromeric and pericentric transcription and transcripts: their intricate relationships, regulation, and functions. *Chromosoma*. 2023 Sep;132(3):211-230.