

The Evolution of Centromeric DNA Sequences

Joshua J Bayes, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

Harmit S Malik, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

Human centromeres are comprised of millions of base pairs of tandemly repeated deoxyribonucleic acid (DNA) sequences. Contrary to the expectation that centromeric sequences would be extremely constrained for centromere function throughout primate evolution, these sequences represent some of the most rapidly evolving sequences in the genome. This rapid evolution in spite of conserved function has been referred to as the 'centromere paradox', and it is hypothesized that an ancient, ongoing genetic conflict is at the heart of this rapid evolution.

Introduction

Centromeres are the chromosomal deoxyribonucleic acid (DNA) regions that are responsible for assembling the kinetochore complex of proteins, which recruits microtubules and mediates chromosomal segregation in all eukaryotes. In spite of stringent evolutionary constraints acting on the process of chromosome segregation, the underlying DNA sequences that define centromeric regions are poorly conserved and evolve quite rapidly even between closely related species. In addition, centromeric proteins also show unexpected signatures of rapid evolution. It has been suggested that an ongoing genetic conflict is at the heart of this rapid evolution (Henikoff *et al.*, 2001).

The Primate Centromeric Satellite: α -Satellite

Centromeres throughout primate lineages are populated by a primate-specific approximately 171-base pair (bp) AT-rich repetitive sequence, α -satellite. Evidence for the role of α -satellite in centromere function came over 15 years ago from experiments that integrated a large stretch of α -satellite DNA into African green monkey chromosomes in cell culture. The integrated α -satellite showed localization of

proteins important for centromeric function and disrupted the normal segregation of chromosomes in these cells, directly implicating it as being the centromeric satellite (Haaf *et al.*, 1992). Given the repetitive nature of centromeric sequences and the difficulty in assembling sequence contigs confidently in the correct order, these sequences are largely missing from human and primate genome sequencing and assembly projects. However, the identification and analysis of DNA contigs that span the 'transition zones' between noncentromeric DNA and centromeric α -satellites on three human chromosomes (8, 17 and X) have provided clues to the evolution of these sequences (Schueler *et al.*, 2005; Rudd *et al.*, 2006). **See also:** [Contig Assembly](#)

Individual α -satellite units (referred to as monomers) are typically arranged in a head-to-tail fashion. These monomers can be further organized into two types of repeat structures (**Figure 1**). At the functional 'core' of centromeric regions, α -satellites are found in a repeat unit that consists of multiple monomers. This multimonomer unit is repeated over and over to make up a higher-order array. Higher-order arrays of α -satellite are the typical sequence organization of centromere regions of humans and can stretch for megabases of DNA that is largely uninterrupted by any kind of insertion or mutation. For example, the repeat unit length of the functional centromere of the human X chromosome is approximately 2 kb, comprised of 12 171-bp monomers of an evolutionarily young α -satellite called DXZ1. All human chromosomes appear to have such a chromosome-specific higher-order satellite DNA array at their centromeres, although the sequence and number of monomers in a repeat unit varies across chromosomes. Comparative sequence analysis of α -satellites from higher-order arrays shows that pairwise comparisons of individual monomers within a single multimeric repeat are only approximately 70–80% identical, while pairwise comparisons of whole multimeric repeats show high homology of approximately 97–99% (**Figure 1**). The high homology of

Advanced article

Article Contents

- Introduction
- The Primate Centromeric Satellite: α -Satellite
- Evolutionary Forces Shaping Centromeric Satellites
- The Foundation of a Functional Kinetochore
- Defining Centromeres: Epigenetics, Not Genetics
- Unexpected Rapid Evolution of Centromere Proteins
- Centromeres in Competition: 'Centromere-drive' and Its Suppression

Online posting date: 30th April 2008

ELS subject area: Evolution and Diversity of Life

How to cite:

Bayes, Joshua J; and, Malik, Harmit S (April 2008) The Evolution of Centromeric DNA Sequences. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.
DOI: 10.1002/9780470015902.a0020827

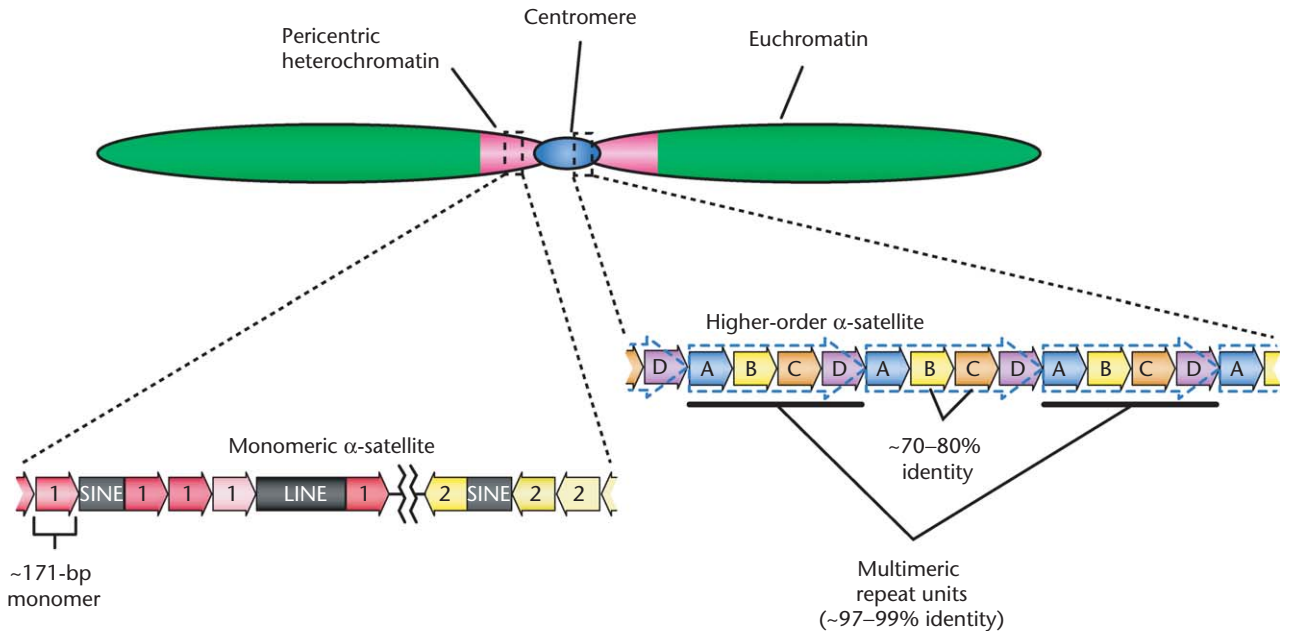


Figure 1 General organization of human centromeric sequences. The centromere is identified as the primary constriction on the chromosome flanked by pericentric heterochromatin. A closer look at the centromere shows that there are two types of α -satellite repeats found at and near human centromeres: monomeric α -satellite repeats (left) and higher-order α -satellite arrays consisting of several monomers repeated as a multimeric unit (right). Monomeric repeats are found in 'domains' in the pericentric regions and display less homogeneity in sequence identity than higher-order α -satellite arrays at the centromere. Frequent insertions of mobile elements disrupt monomeric α -satellite domains while higher-order α -satellite arrays can span megabases of DNA largely uninterrupted.

the higher-order α -satellite repeats within a centromere is consistent with their current function as the centromeric DNA where selection may favour homogeneity to retain centromere function.

Flanking the centromere 'proper', α -satellites can be found in monomeric units. These α -satellites are considered pericentric; whereas they may serve important roles in chromosome segregation, they do not recruit centromeric or kinetochore proteins and are less homogenous than higher-order arrays. For example, pairwise comparisons of pericentric monomers range in identity from approximately 60 to 75% (Figure 1). Since pericentric monomers of α -satellites are no longer necessarily critical for centromere function, little selective constraint acts on these sequences and they are subject to spontaneous mutational events and insertions by mobile elements. Analysis of the pericentric regions of human chromosomes has revealed that monomeric units are organized into relatively small contiguous domains, wherein monomeric units within a domain are more closely related and believed to have been homogenized in sequence independent of monomers in other domains. These separate pericentric domains reveal a fascinating palaeontological record of satellite arrays that may have represented the functional centromeric regions in ancestral primates. As new sequences arose and became the new functional centromeric sequence, these ancestral centromeric repeats were pushed out to the pericentric regions. At least five such events are evident in the human X pericentric region, with the most ancient domains also being

the most distal from the current centromere (Schueler *et al.*, 2005). Such data supports the notion that the higher-order repeat units seen in human centromeric regions are a relatively recent invention, as presumed primordial centromeres lack a discernible higher-order structure.

Evolutionary Forces Shaping Centromeric Satellites

Three evolutionary forces predominantly shape centromeric DNA: recombination, mutation/transposition and selection. One can surmise that recombination plays a central role in centromeric DNA evolution based on the high homogeneity of the centromeric higher-order α -satellite arrays. It is hypothesized that this could be due to unequal sister chromatid exchange and gene conversion (Smith, 1976), which can eliminate sequence heterogeneity as well as remove any insertions by mobile elements, essentially purifying the centromeric repeats of any variation. Unequal sister chromatid exchange offers an explanation for the observation that higher-order α -satellite arrays on nonhomologous chromosomes are substantially different, ranging from 2 to 35 monomers per repeat unit, and only about 70–80% identical at the sequence level. This type of evolution is referred to as 'concerted evolution' where each α -satellite array homogenizes independently for each chromosome. **See also:** [Concerted Evolution](#)

The organization of the pericentric monomers and centromeric higher-order repeats might suggest a model in which higher-order repeats arose via the organization of adjacent monomeric satellites i.e. a purely recombinational process. However, recent studies have demonstrated that this might not be the case. Instead, a higher-order array architecture may have arisen at one centromere recently in primate evolution (around the gorilla–orangutan split) and spread to other chromosomes via transposition (Schueler and Sullivan, 2006). Subsequently, unequal exchanges or gene conversions amplified the higher-order arrays and resulted in the species-specific higher-order centromeric arrays that are predominantly observed at different human chromosomes. In addition, mutation has introduced new variants that have swept to fixation in some centromeres. For instance, the DXZ1 satellite that comprises the entire human X centromeric array is evolutionarily young, believed to be younger than 8 million years old (Schueler *et al.*, 2001). The finding that similar arrays are not always found on orthologous chromosomes in different primate species lends support to the idea that such arrays can be ‘seeded’ by transposition events from nonhomologous chromosomes.

There is no a priori expectation that satellite repeats should evolve faster than nonrepetitive DNA in the absence of any biases introduced by selection. This is because mutations in any particular satellite repeat (introduced with a mutation rate, μ) have a probability of fixation that is proportional to their initial incidence ($1/2N$, where $2N$ are the number of repeat units in arrays on both homologous chromosomes). Thus, the overall likelihood for any mutation spreading to fixation summed over the entire array equals $2N$ times $\mu/2N$, which equals the mutation rate (μ) for nonrepetitive DNA. In this context, we can evaluate what role selection plays in the evolution of centromeric repeats. If stringent constraint were acting on centromeric repeats (as one might expect), then the probability of fixation of mutations in the array is impeded, and so the overall array mutation rate is much less than μ . If, however, the rate of fixation of mutations is somehow accelerated, then the effective array mutation rate will exceed μ . The pericentric monomers provide a good yardstick for this comparison because they are presumed to be selectively neutral (or nearly so). Comparison of the monomeric units and centromeric higher-order array units from orthologous chromosomes (e.g. chimp versus human) leads to the surprising finding that the centromeric arrays from different species are more divergent than the pericentric units (Rudd *et al.*, 2006). These findings are counterintuitive because the centromeric α -satellite array is the functional centromere and is under stringent selective constraint, whereas the pericentric α -satellites are not. Thus, one is left with the paradoxical observation that the satellite units that are most constrained *within* a species have evolved most rapidly *between* species. It is this paradoxical observation that leads to the idea that some selective force must actively drive the rapid fixation of mutations at centromeric satellites by imposing a bias in favour of retaining mutations, thereby increasing the mutation rates of the

entire array. It has been suggested that this selective force may be due to the selfish advantage conferred to centromeres during female meiosis, or ‘centromere-drive’ (Henikoff *et al.*, 2001) (described later).

The Foundation of a Functional Kinetochores

The vital function of the centromeric DNA repeats is to sponsor the assembly of a large complex of proteins referred to as the kinetochore, which provides the physical link between chromosomal DNA and the microtubules. It is these microtubules that will ‘pull apart’ sister chromosomes to ensure equal segregation. In addition to binding microtubules, the kinetochore hosts a bevy of proteins that monitor microtubule attachment and proper tension across the centromere. Unattached microtubules or unequal tension will elicit a cascade of signalling events initiating from the kinetochore, halting cell division until the problem is corrected. Thus, the kinetochore complex is endowed with the sophisticated function of ensuring the faithful segregation of chromosomes at each cell division. More than 60 kinetochore proteins have been identified from the budding yeast *Saccharomyces cerevisiae*, and many of these have human orthologues. Three human kinetochore ‘foundation proteins’ that help define the centromere and directly bind centromeric DNA will be discussed. **See also:** [Kinetochore: Structure, Function and Evolution](#)

CENP-A is a histone H3 variant whose essential role is to replace H3 in centromeric nucleosomes (**Figure 2**). Homologues of *CENP-A* (also referred to as CenH3s, or centromeric histones) have been identified in all eukaryotes examined in detail; they are critically important for centromere function. *CENP-A* establishes the epigenetic nature of the centromere by packaging centromeric DNA into a unique chromatin structure that differs from the rest of the genome. All active centromeres are characterized by the presence of *CENP-A*. Very early on after the identification of the α -satellite, it was remarked that the length of the α -satellite monomer was almost exactly what would be expected of a DNA bound by a single nucleosome. Thus, the original model proposed (that is still prevalent) has been that the fundamental unit of centromeres is a single centromeric nucleosome (Palmer *et al.*, 1991).

CENP-C is an essential gene encoding a constituent protein of the kinetochore foundation. *CENP-C* localization to centromeric regions is dependent on *CENP-A* and it has been suggested that *CENP-C* may specifically interact with the altered chromatin structure induced by *CENP-A* (**Figure 2**). Thus, while *CENP-C* centromeric DNA localization domains have been mapped to the middle regions of the protein, no particular binding sequence has been identified. More recently, *CENP-C* has been shown to possess ribonucleic acid (RNA)-binding abilities that appear to be more specific (Wong *et al.*, 2007). Whether these DNA- and RNA-binding activities directly contribute to *CENP-C* localizing to the centromeric region is still unclear.

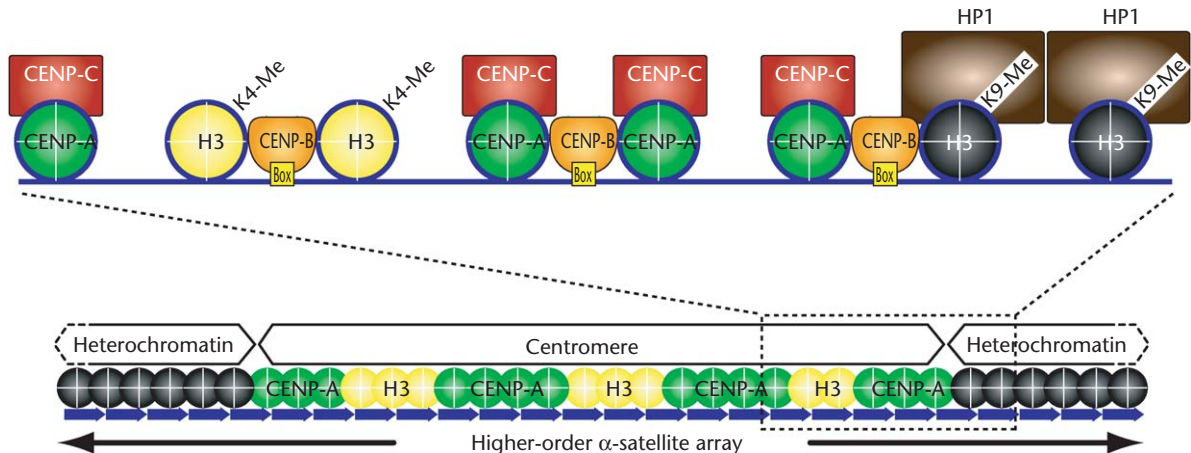


Figure 2 Higher-order α -satellite arrays host both heterochromatin and centromeric domains. The centromere is packaged into a discontinuous array of nucleosomes containing CENP-A or canonical histone H3 flanked on either side by heterochromatin. The localization of 'foundation' proteins is shown enlarged in a close-up view. CENP-B associates throughout the array bound to its CENP-B box, while CENP-A and CENP-C are found only at a subset of the region bound by CENP-B. Centromeric regions packaged into H3-containing nucleosomes show marks associated with 'active' chromatin typical of euchromatin (K4-Me). In contrast, neighbouring heterochromatin is packaged into H3-containing nucleosomes with a different mark (K9-Me) and localize the prototypical heterochromatin protein, HP1.

CENP-B protein is unique among kinetochore proteins in that it binds a specific 17-bp sequence (the CENP-B box) found within a subset of α -satellite monomers (Figure 2). The exact role of CENP-B proteins in centromere function is still not completely elucidated. It is clear that unlike CENP-A or CENP-C, CENP-B is not essential for mitotic function (as demonstrated by the viability of knockout mice). CENP-B can be present at both the inactive and active centromeres of a dicentric chromosome, suggesting that its localization is not simply associated with centromere function. In addition, centromeric satellites on some chromosomes lack CENP-B boxes (the Y-chromosome for instance) yet form a functional centromere. Moreover, in contrast to the conservation of CENP-B throughout eukaryotes, the CENP-B box has not been conserved as widely even among primate centromeric arrays. However, CENP-B binding has been demonstrated to increase the efficiency of CENP-A chromatin formation on human artificial chromosomes. Despite its nonessential function in humans, homologues of CENP-B have also been identified throughout various phyla with the fission yeast *Saccharomyces pombe* containing three homologues where these proteins are essential to cell viability.

The foundation proteins are organized differently on human centromeric DNA (Figure 2). CENP-B is present all across the higher-order repeat structure, but CENP-A and CENP-C are only found in some blocks of repeat units, interspersed with blocks of canonical (H3-containing) nucleosomes that possess histone modifications more characteristic of euchromatic than heterochromatic regions. It is believed that these blocks of CENP-A self-organize to present one combined 'surface' to organize the rest of the kinetochore proteins, which will serve as the microtubule attachment sites. The pericentric monomers that flank the centromeric repeat units are believed to be largely devoid

of the foundation proteins, and are instead packaged with canonical nucleosomes that possess histone modifications characteristic of heterochromatin and bound by heterochromatin proteins like HP1. This heterochromatin boundary is important both to define the boundaries of centromeric domains, as well as to recruit cohesin proteins that hold the sister chromatids together until they are finally ready to separate during anaphase of the cell cycle. Work in fission yeast has determined that chromosome segregation occurs properly only when a 'boundary' between heterochromatin and the centromeric domain is present, preventing the encroachment of heterochromatin into the centromere, and vice versa. In human centromeres, such a boundary has not yet been localized, yet it is evident that such a boundary must exist. For example, recent studies of human artificial chromosomes have shown that an integrated array of α -satellite of only 10 kb onto the chromosomal arm in human cell lines is heterochromatic and does not recruit CENP-A (Okamoto *et al.*, 2007). However, if heterochromatin formation is inhibited then CENP-A is present at this chunk of 10 kb α -satellite. Therefore, the particular location and the amount of CENP-A deposition in a sea of repetitive DNA may be governed by a balance between CENP-A chromatin and heterochromatin, as well as additional factors discussed later. **See also:** [Heterochromatin: Constitutive](#)

Defining Centromeres: Epigenetics, Not Genetics

From the perspective of centromeric DNA evolution, the most important question is how the location of the centromere is determined, or, in other words, how does the kinetochore 'know' where to assemble? Because the

primary sequences of centromeres vary so greatly among species, and yet general conservation of kinetochore proteins is evidenced by the identification of homologues in diverse taxa, an epigenetic, sequence-independent mechanism, of centromere determination is widely favoured in the field (Sullivan *et al.*, 2001). This is in contrast to a genetic mechanism where the centromeric sequences are directly recognized by kinetochore proteins (akin to a transcription factor recognizing a binding site within an enhancer). Several lines of evidence have suggested that a 'genetic' definition of centromere function in humans is not an adequate model. Evidence rules out α -satellite as the primary sequence essential for CENP-A binding since human neocentromeres that lack α -satellite sequences can nonetheless localize CENP-A. There are also many examples of CENP-A spreading to noncentromeric sequences in studies of human artificial chromosomes.

It has been suggested that the basic unit of centromere inheritance is the CENP-A nucleosome. Such centromeric nucleosomes assort randomly to daughter chromosomes following DNA replication, and may catalyse the deposition of new CENP-A nucleosomes in 'nucleosomal gaps', thereby ensuring that satellite repeats that served centromeric function will continue to do so even after cell division. Such epigenetic inheritance is clearly visualized in the case of the dicentric chromosomes, which have to inactivate one of their two centromeres to avoid chromosome loss. The inactivated centromere shows no CENP-A localization, whereas the active centromere has CENP-A localization. Following cell division, the 'active' centromere continues to recruit CENP-A and maintain centromeric function while the inactive centromere remains quiescent, despite both possessing α -satellites.

What is not in dispute, however, is the fact that α -satellites can 'seed' the formation of centromeres, even apparently *de novo*. For instance, large segments of DNA containing the DXZ1 α -satellite higher-order array from the X-chromosome recruit centromeric proteins and can be stably inherited in cell culture. Furthermore, recent studies in human artificial chromosome construction show that tens of kilobases of α -satellite correlate with efficient retention of human artificial chromosomes in human cell culture (the presence of CENP-B boxes increased mitotic fidelity) (Okamoto *et al.*, 2007). This has led to the suggestion of the possibility that α -satellite may itself determine the efficiency of centromere formation. The ability of α -satellite to more efficiently form centromeres could be a consequence of the unique chromatin structure needed for proper centromere function. Structural studies have shown that centromeric nucleosomes impose a specialized chromatin structure that is not found elsewhere in the genome. Overexpression of CENP-A leads to the deposition of centromeric nucleosomes outside the centromeric regions (for instance, in euchromatin). However, these mislocalized CENP-A nucleosomes are quickly eliminated by proteolysis everywhere except the centromere. Thus, it has been proposed that it may not be the sequence of the α -satellite per se, but its ability to adopt

a unique chromatin structure on which the kinetochore complex can assemble.

Thus, an epigenetic mechanism that confers a special ability on α -satellites to recruit the CENP-A nucleosome can explain several puzzling features of centromere maintenance. However, even this mechanism still leaves unexplained the finding that α -satellite sequences have diverged more rapidly than even neutral sites. These evolutionary findings suggest that some aspect of α -satellite sequence is directly playing a role in centromeric protein recruitment, although the exact means by which this is accomplished remains to be determined. **See also:** [Epigenetic Factors and Chromosome Organization](#)

Unexpected Rapid Evolution of Centromere Proteins

Because of the repetitive nature of centromeric sequences and the difficulty of manipulating such sequences, one way to understand the importance of these sequences and to gain insight into centromere biology is to 'back away' and look at the evolutionary pressures shaping centromere proteins. The advantages to such an approach is that types of selection on genes that encode the foundation proteins are more easily discernable than evolutionary studies on noncoding DNA. This approach also avoids the difficulties associated with the massive sequencing and assembly efforts required to understand centromeric DNA evolution.

Once again, such analyses that compare rates of evolution of centromeric proteins (changes at nonsynonymous versus synonymous sites) reveal an unexpected finding. Since these centromeric proteins are absolutely critical for chromosome segregation, one might have expected that these proteins should be subject to stringent constraint and intolerant of amino acid changes ('purifying selection'). Instead, there is strong evidence for 'positive selection' (excess amino acid changes) in centromeric proteins across many animal and plant lineages. Positive selection refers to selection that favours protein innovation, with the classical example being the type of selection observed in immune proteins that evolve rapidly to recognize a changing landscape of pathogenic epitopes. While not surprising that genes encoding proteins with immune functions would be under positive selection driven by host-virus conflicts, it is completely unexpected to discover positive selection among genes encoding proteins involved in centromere functions. This suggests that a conflict between centromere proteins and centromeric DNA exists at the centromere, which drives the rapid evolution of both centromeric DNA as well as centromeric proteins. **See also:** [Intragenomic Conflict; Synonymous and Nonsynonymous Rates](#)

Centromeres in Competition: 'Centromere-drive' and Its Suppression

Conflicts between centromeres and their binding proteins seem counterintuitive since proper centromere function is so critical for survival. How could such a conflict exist? Centromeres mediate two types of chromosome segregations: mitosis and meiosis. A successful mitosis results in two daughter cells that are genetically identical – each chromosome has the same evolutionary fate. The products of meiosis, however, are haploid and not genetically identical, leading to the possibility of genetic conflict acting at meiosis (meiotic drive), or shortly after (post-meiotic dysfunction). Furthermore, whereas male meiosis produces four haploid products, which will all become sperm, female meiosis will discard three of its four haploid products, keeping only one to become the egg pronucleus. Female meiosis therefore provides an extraordinary opportunity for any element to gain an evolutionary advantage by insuring its inclusion in the egg. Centromeres are a prime

candidate for taking advantage of this meiosis since they are the primary sites of microtubule attachment and could play a role in directly orienting chromosomes for preferential inclusion into the egg. **See also:** [Meiosis](#)

Indeed, transmission biases of chromosome rearrangements involving the centromeric regions have been observed. Robertsonian chromosomes refer to the fusion of two *acrocentric* chromosomes (chromosomes with centromeres near the end of the chromosomes) to now become a *metacentric* (where the centromere is found near the middle of the chromosome). Robertsonian fusion chromosomes are found at a frequency of about 1 in every 1000 individuals. They are genetically identical to the acrocentric chromosomes from which they are derived, and therefore manifest no somatic defects. By looking at the genotype of the children from carriers of a Robertsonian fusion, it was determined that the fusion chromosome was transmitted approximately 58% of the time when the carrier was the mother but at the expected rate of 50% when the father was the carrier. It was concluded that inherent asymmetries associated with female meiosis explains this observation. Because of this preferential transmission, this type of

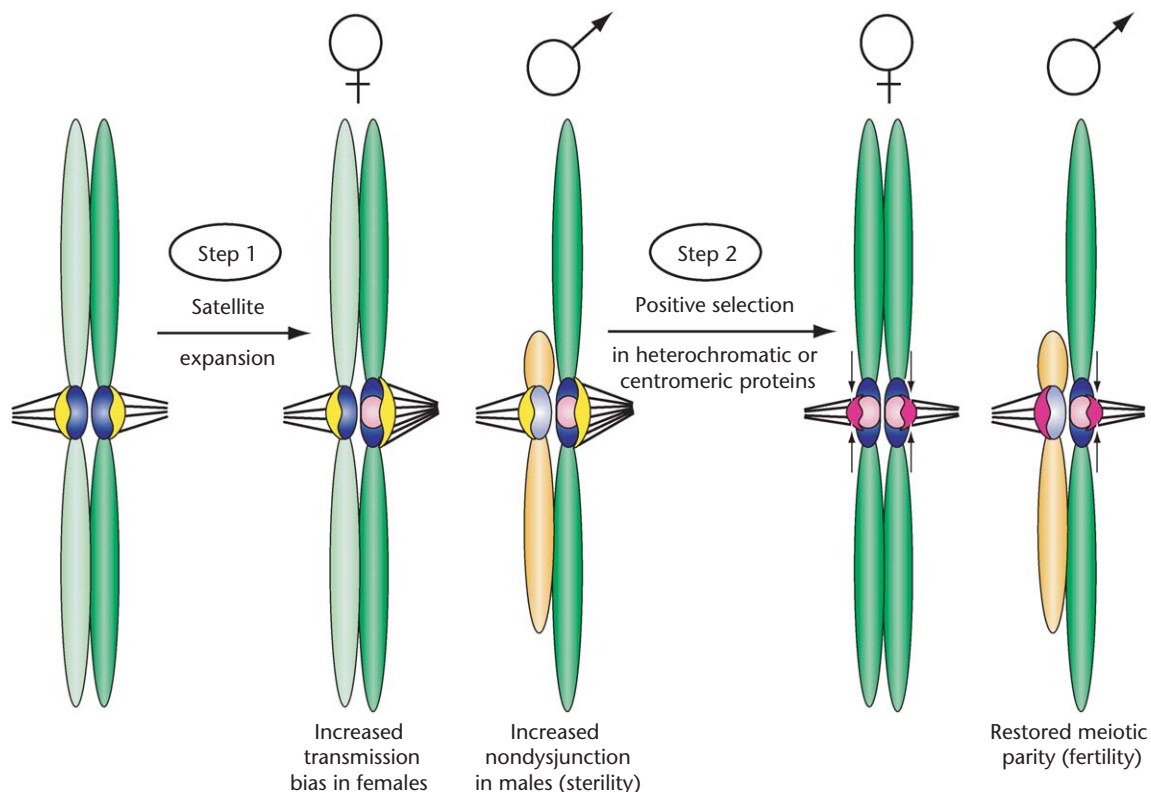


Figure 3 'Centromere-drive' model. The sex chromosomes provide an exaggerated example of centromere-drive since the X and Y chromosomes do not recombine and the satellites populating each respective centromere are different (shown as different colours). Step 1, the centromere variant has a selective advantage by virtue of its ability to attract more microtubules and gain a favourable position in the asymmetric female meiosis. However, this centromere variant results in unequal tension across the centromeres of the paired sex chromosomes in male meiosis. Unequal tension can potentially increase the rate of nondysjunction in males, resulting in sterility. Step 2, sterility effects in the male will provide strong selection on any allele that restores meiotic parity and male fertility. Repeated bouts of drive and suppression will be observed as positive selection among centromere or heterochromatin proteins. Shown here is an example of selection favouring alleles that no longer recognize the 'old' centromeric satellite (on the X) resulting in constriction of the centromere to only the new satellite variant and restoring relatively equal microtubule binding (while the Y centromeric satellite still retains its binding ability).

phenomenon has been predicted to have a profound effect on chromosome evolution in humans (Pardo-Manuel de Villena and Sapienza, 2001). For instance, chromosome 2 is derived from the fusion of two acrocentric chromosomes sometime after the human–chimpanzee species split and is now metacentric in humans.

Given these observations, a hypothesis has been proposed to explain the expansion and/or fixation of centromeric satellite variants (Figure 3) (Henikoff *et al.*, 2001). Coined the ‘centromere-drive’ hypothesis, it posits that any sequence variation at the centromere that allows a centromere to ‘win’ in female meiosis will be greatly favoured and fixed rapidly in the population. The advantage gained by the centromere is predicted to be a result of preferential positioning in female meiosis presumably by its interactions with microtubules (e.g. by binding more microtubules). Novel variation in α -satellite sequence, organization, or even just the amount of α -satellite potentially provides an increased opportunity for CENP-A incorporation and a greater capability for microtubule binding. This model provides a tantalizing explanation for the unexpected rapid evolution of centromeric repeat units as well as the shift from monomeric to higher-order α -satellites in primates. Additionally, centromere-drive may also help explain the intriguing finding from human polymorphism data that centromeric regions have been at the centre of repeated adaptive sweeps in recent human history (Williamson *et al.*, 2007).

Centromere-drive, while advantageous to the ‘selfish’ driving centromere, could be strongly disadvantageous to the organism as it could cause meiotic defects particularly in the male. A male heterozygous for two different centromere variants (for instance, one that recruits more microtubules than the other) will potentially result in unequal tension between the centromeres. Male meiosis has been shown to be particularly sensitive to differences in the amount of tension placed on each centromere and results in a pause in meiosis until tension is equalized. Thus, there could be a striking cost of centromere-drive on male fertility. For example, a significant fraction of male carriers of Robertsonian fusions have some degree of sterility associated with this karyotype, despite being favoured in females. Left unchecked, this could lead to a significant impact on the species.

Therefore, centromere-drive and the rapid gains of the selfish driving centromere in the population, in turn, places stringent selective pressure on the rest of the genome to find a means by which it can restore meiotic parity and prevent male sterility. It has been proposed that this could be accomplished by the alteration of DNA-binding preferences among centromeric or heterochromatin proteins to no longer favour older satellites, for instance, and equalize tension across the centromeres. Such changes would be selectively advantageous and could lead to rapid fixation of alleles that restore male fertility. Therefore, positive selection in centromeric proteins is hypothesized to be a result of conflicts between the protein and the driving centromeric satellite as evolution continually favours mutations in these

proteins that suppress centromere-drive and alleviate sterility effects in the male.

Centromere-drive provides an intriguing explanation of the rapid evolution of centromeric DNA, while the suppression of centromere-drive provides the selective rationale for the unexpected positive selection identified in essential proteins required for centromere function. Despite having to work in concert for one of the most fundamental processes in eukaryotic biology, centromeres and centromere-binding proteins are nonetheless locked in a genetic conflict, not unlike that seen between a selfish genetic element and the host genome.

References

- Haaf T, Warburton P and Willard H (1992) Integration of human alpha-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. *Cell* **70**(4): 681–696.
- Henikoff S, Ahmad K and Malik H (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**(5532): 1098–1102.
- Okamoto Y, Nakano M, Ohzeki J *et al.* (2007) A minimal CENP-A core is required for nucleation and maintenance of a functional human centromere. *EMBO Journal* **26**(5): 1279–1291.
- Palmer DK, O’Day K, Trong HL *et al.* (1991) Purification of the centromere-specific protein CENP-A and demonstration that it is a distinctive histone. *Proceedings of the National Academy of Sciences of the USA* **88**(9): 3734–3738.
- Pardo-Manuel de Villena F and Sapienza C (2001) Female meiosis drives karyotypic evolution in mammals. *Genetics* **159**(3): 1179–1189.
- Rudd M, Wray G and Willard H (2006) The evolutionary dynamics of alpha-satellite. *Genome Research* **16**(1): 88–96.
- Schueler M, Dunn J, Bird C *et al.* (2005) Progressive proximal expansion of the primate X chromosome centromere. *Proceedings of the National Academy of Sciences of the USA* **102**(30): 10563–10568.
- Schueler M, Higgins A, Rudd M *et al.* (2001) Genomic and genetic definition of a functional human centromere. *Science* **294**(5540): 109–115.
- Schueler M and Sullivan B (2006) Structural and functional dynamics of human centromeric chromatin. *Annual Reviews of Genomics and Human Genetics* **7**: 301–313.
- Smith G (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* **191**(4227): 528–535.
- Sullivan BA, Blower MD and Karpen GH (2001) Determining centromere identity: cyclical stories and forking paths. *Nature Reviews. Genetics* **2**(8): 584–596.
- Williamson SH, Hubisz MJ, Clark AG *et al.* (2007) Localizing recent adaptive evolution in the human genome. *PLoS Genetics* **3**(6): e90.
- Wong LH, Brettingham-Moore KH, Chan L *et al.* (2007) Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Research* **17**(8): 1146–1160.

Further Reading

- Alkan C, Ventura M, Archidiacono N *et al.* (2007) Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Computational Biology* **3**(9): 1807–1818.
- Amor DJ and Choo KH (2002) Neocentromeres: role in human disease, evolution, and centromere study. *American Journal of Human Genetics* **71**(4): 695–714.
- Amor DJ, Kalitsis P, Sumer H *et al.* (2004) Building the centromere: from foundation proteins to 3D organization. *Trends in Cell Biology* **14**(7): 359–368.
- Choo KHA (1997) *The Centromere*. New York: Oxford University Press.
- Karpen GH and Allshire RC (1997) The case for epigenetic effects on centromere identity and function. *Trends in Genetics* **13**(12): 489–496.
- Malik HS and Henikoff S (2002) Conflict begets complexity: the evolution of centromeres. *Current Opinions in Genetics and Development* **12**(6): 711–718.
- Warburton PE and Willard HF (1996) Evolution of centromeric α -satellite DNA: molecular organization within and between human and primate chromosomes. In Jackson M, Strachan T and Dover G (eds) *Human Genome Evolution*, pp. 121–145. Oxford: BIOS Scientific Publishers.
- Willard HF (1985) Chromosome-specific organization of human alpha satellite DNA. *American Journal of Human Genetics* **37**(3): 524–532.
- Willard HF (1991) Evolution of alpha satellite. *Current Opinions in Genetics Development* **1**(4): 509–514.