

Title

The evolutionary life cycle of the resilient centromere

Paul Kalitsis^{1,2*} and K.H. Andy Choo^{1,2}

¹Murdoch Childrens Research Institute, Royal Children's Hospital, Melbourne, Victoria 3052, Australia.

²Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Victoria 3052, Australia.

*Corresponding author

Abstract

The centromere is a chromosomal structure that is essential for the accurate segregation of replicated eukaryotic chromosomes to daughter cells. In most centromeres, the underlying DNA is principally made up of repetitive DNA elements, such as tandemly-repeated satellite DNA and retrotransposable elements. Paradoxically, for such an essential genomic region the DNA is rapidly evolving both within and between species. In this review, we show that the centromere locus is a resilient structure that can undergo evolutionary cycles of birth, growth, maturity, death and resurrection. The birth phase is highlighted by examples in humans and other organisms where centromere DNA deletions or chromosome rearrangements can trigger the epigenetic assembly of neocentromeres onto genomic sites without typical features of centromere DNA. In addition, functional centromeres can be generated in the laboratory using various methodologies. Recent mapping of the foundation centromere mark, the histone H3 variant, CENP-A, onto near-complete genomes has uncovered examples of new centromeres which have not accumulated centromere repeat DNA. During the growth period of the centromere, repeat DNA begins to appear at some, but not all, loci. The maturity stage is characterised by centromere repeat accumulation, expansions and contractions, and the rapid evolution of the centromere DNA between chromosomes of the same species and between species. This stage provides inherent centromere stability, facilitated by repression of gene activity and meiotic recombination at and around the centromeres. Death to a centromere can result from genomic instability precipitating rearrangements, deletions, accumulation of mutations and the loss of essential centromere binding proteins. Surprisingly, ancestral centromeres can undergo resurrection either in the field or in the laboratory, via as yet poorly understood mechanisms. The underlying principle for the preservation of a centromeric evolutionary life cycle is to provide resilience and perpetuity for the all-important structure and function of the centromere.

You can threaten me, change me, or kill me, but you can't get rid of me. I am here to stay.

- The resilient and all-important centromere

Introduction

The centromere is an essential chromosomal structure that has multiple functional roles for the accurate segregation of replicated chromosomes to daughter cells. These roles include: genetic/epigenetic marking and assembly of a proteinaceous kinetochore complex during the cell cycle, mitotic checkpoint, sister chromatid cohesion and release, chromosome movement and cytokinesis (Santaguida and Musacchio 2009; Allshire and Karpen 2008; Perpelescu and Fukagawa 2011). A defect in any one of these functions can result in aneuploid cells with gained or lost chromosomes. In mammals, aneuploidy is a major contributor to birth disorders, spontaneous abortions, and infertility (Hassold and Hunt 2001), and is often associated with cancer due to the loss of tumour-suppressor genes or gain of oncogenes (Weaver and Cleveland 2007; Thompson et al. 2010).

In eukaryotes, centromeres are classified into two groups, simple and complex. At the core of each type of centromere is the primary nucleosomal mark, the histone H3 variant CENP-A or CenH3, which replaces the canonical histone H3 upon which downstream centromere proteins bind to and build a mature kinetochore (Black and Cleveland 2011). Simple, “point”, or genetic, centromeres are strictly dependent on specific DNA motifs for the binding of centromere proteins, as are present in single-celled fungi in the *Saccharomycetaceae* family such as *Saccharomyces cerevisiae* and *Kluyveromyces lactis* (Malik and Henikoff 2009). Complex, or “regional”, centromeres found in single and multi-celled organisms consist of long stretches (from tens of, to a few thousand kilobase pairs) of tandemly repeated satellite DNA and/or retrotransposable elements, which are sufficient but not essential for the binding of centromere proteins. This flexibility in the binding of centromere

proteins to DNA highlights an epigenetic nature of the complex centromeres that affords them the ability to be turned on or off as a means of adaptation to genomic rearrangements in cells and driving speciation.

Here, we review evidence from comparative genome analyses, laboratory data and human cytogenetic findings that point to the centromere having an inherent capability to undergo a process of birth, growth, maturity, death and even resurrection, in a so-called “centromere life cycle” (Fig. 1). Elements of such a life cycle could be detected over a relatively short period of time, even over a few generations, while the occurrence of other elements required a longer evolutionary time span.

Birth

Evolutionary birth of centromeres

A primitive centromere-like system appears to have simultaneously arisen with the first appearances of single-celled life forms in bacteria and Archaea. A chromosome partitioning system had evolved in concert with an increase in the amount of cellular genetic material, whose inheritance to future generations had necessitated such a DNA partitioning mechanism. The chromosome/plasmid partitioning systems in these simple organisms are very rudimentary and are based upon three components: (i) a DNA element that is recognised by, (ii) a trans-acting protein that links up, (iii) the polymer-based filaments that push the chromosomes to opposite poles (Malik and Henikoff 2009).

As cells became more complex, accompanied by an increase in DNA amount, the presence of distinct cell cycle phases and the establishment of RNAi/heterochromatin machinery in most eukaryotes, so did centromeres. However, some single-celled fungal species in the *Saccharomycetaceae* family could possibly have lost their RNAi/heterochromatin machinery and have reverted back to sequence-dependent centromeres such as the small point centromere of *Saccharomyces cerevisiae* and *Kluyveromyces lactis* (Heus et al. 1993; Meraldi et al. 2006; Malik

and Henikoff 2009). In contrast, other single-celled fungi with centromeric heterochromatin, such as *Schizosaccharomyces pombe*, and most multi-cellular eukaryotes have larger centromeres ranging in length from tens of kilobases to several megabases. These centromeres mainly consist of either tandemly repeated arrays of DNA or retrotransposable elements, or a mixture of both. What also made these repeat-rich centromeres different to the prokaryotic and point centromeres was the loss of a strict DNA sequence dependency; in other words, these larger/complex centromeres had shifted from a strictly genetic state to a genetic-come-epigenetic state.

Neocentromeres

The epigenetic nature of centromeres was nicely illustrated by the discovery of neocentromeres both naturally and in the laboratory. Neocentromeres are centromeres that appear in hitherto non-centromeric chromosomal regions that are usually devoid of centromeric repeat DNA. The new centromeric sites can be stably inherited through future cell divisions, across many generations, and spanning millions of years (see Evolutionary New Centromeres section).

In humans, neocentromeres have been reported in over 90 cases where a chromosomal rearrangement has triggered the ‘birth’ of a new centromere (Warburton 2004; Marshall et al. 2008a). The most common observed cause of neocentromere formation is a chromosome rearrangement event producing an acentric or centromerically compromised (such as through partial deletion) fragment that is rescued by the new centromere. In most of these cases, the neocentromere is positioned in a gene-poor euchromatic region with no discernible sequence preference. Some clustering of reported neocentromere sites towards the ends of chromosomes has been reported, caused by the frequent loss of interstitial regions near to the deleted original centromere sites. From the available reported neocentromere cases there is no obvious direct association between the appearance of a neocentromere and specific disease; however most cases are associated with developmental disorders due perhaps to bias in the ascertainment of cases presenting at the genetic clinics.

Assisted centromere birthing

Historically researchers have mapped centromeres either by genetic or physical means. Once the region or candidate sequence has been identified then the naked DNA of the corresponding sequence can be reintroduced back into the cell as an artificial chromosome construct to elicit its centromere function. Another path used to generate or activate a centromere involves forcing genomic rearrangements or centromere protein overexpression to trigger the seeding of a new centromere in a previously non-centromeric site. A number of specific strategies have been used to effectively facilitate *de novo* centromere formation (Table 1):

(i) Production of acentric chromosomal fragments

Gamma irradiation had been used as an agent to fragment chromosomes in the fly, *Drosophila*. These experiments produced acentric fragments that, when characterised, did not contain the usual centromeric repeat DNA but were able to attract essential centromere proteins to instate functional chromosome segregation activity (Williams et al. 1998). In a separate experimental strategy, FLP-mediated recombination was used to release a distal block of heterochromatin from a *Drosophila* chromosome. The released, otherwise acentric, fragment formed a new centromere and was able to segregate faithfully to daughter cells (Platero et al. 1999).

In the larger chromosomes of plants, such as barley and wheat, Nasuda and colleagues induced chromosome truncation of a single barley chromosome in a wheat background (Nasuda et al. 2005). The native, repeat-rich centromere was deleted. Fluorescence *in situ* hybridisation (FISH) analysis showed that the derivative chromosome lacked any known centromeric repeats but contained a new centromere carrying essential centromere proteins.

(ii) Eviction of endogenous centromere

In the fission yeast, the Cre-loxP system was used to excise the native centromere (Ishii et al. 2008). This forced centromere eviction led to either the activation of neocentromere-containing

chromosomes or the rescue of the acentric fragment via telomere-telomere fusion. Neocentromeres were found to preferentially form at subtelomeric regions of the chromosome.

Unlike fission yeast, the small regional centromeres of the pathogenic fungus *Candida albicans* contain no pericentric heterochromatin (Ketel et al. 2009). To test whether neocentromeres could be produced in this organism, the chromosome V centromere was replaced with the URA3 marker gene. Neocentromeres were able to form at multiple locations along chromosome V with no obvious sequence preference, however direct and inverted repeats were located nearby (Marshall and Choo 2009).

(iii) Insertion of centromere DNA into cells

The introduction of putative centromere DNA sequences into cells has been used as an assay to define the minimal region required for *de novo* centromere activity in a multitude of organisms. Early experiments in budding yeast showed that the centromere has strict sequence dependency for full chromosome segregation activity (Fitzgerald-Hayes et al. 1982). Similar studies in other point-centromere species from the same phylogenetic group such as *Kluyveromyces lactis*, *Candida maltosa* and *Candida glabrata* showed that cen-DNA plasmids were mitotically stable but, surprisingly, were not able to function in *S. cerevisiae* even though the centromeres shared similar sequence structure (Heus et al. 1990; Ohkuma et al. 1995; Kitada et al. 1996; Stoyan and Carbon 2004).

The activation of a fully functional centromere from *in vitro* DNA constructs with regional centromeres has proved to be more challenging. For example, the closest characterised regional/epigenetic centromere species to the point-centromere group is *Candida albicans*; this species resides within the *Saccharomycotina* sub-phylum but in a neighbouring clade to the point-centromere species (Wang et al. 2009). Attempts at transforming plasmid vectors carrying centromere-spanning DNA fragments have not shown any mitotic stability or attraction of CENP-A

protein (Baum et al. 2006). It may be that the cen-DNA plasmids need to be suitably epigenetically primed before they can elicit centromere activity.

By contrast, transformation of yeast artificial chromosomes (YACs) containing the larger regional centromeres (40 to 100 kb) of the fission yeast, *S. pombe*, proved to be more successful (Hahnenberger et al. 1989). It was shown that the flanking repetitive regions were needed for full chromosome stability of the artificial chromosomes, but some of the deleted centromere DNA YAC clones were able to confer full mitotic stability. It is speculated that these abbreviated constructs may require several cell divisions to acquire the full centromeric epigenetic marks for mitotic stability (Steiner and Clarke 1994).

In other model organisms, such as the filamentous fungus, *Neurospora* and the fruit fly, *Drosophila*, researchers have had no success in activating centromere activity from naked DNA constructs. Reasons for this may include the constructs not harbouring adequate amounts of DNA from the centromeric locus, or the cell lines of choice lacking the relevant epigenetic factors for *de novo* centromere assembly.

(iv) Cen DNA transfection (de novo formation of centromeres in mammalian cells)

During the last two decades, a great deal of effort has been put into the induction of *de novo* centromere activity and the creation of artificial chromosomes in mammalian, in particular human and mouse cell lines. Early mammalian transfection experiments attempted to replicate the centromere DNA assays that had been successfully used in fungi. The integration of human alpha satellite DNA into mammalian cell lines attracted centromere proteins and exhibited centromere function but not full mitotic stability (Haaf et al. 1992; Larin et al. 1994). These studies were subsequently followed by a more refined human artificial chromosome construct complete with centromere, telomere and intervening genomic DNA for transfection into human cells (Harrington et al. 1997). The centromere DNA sequences within such artificial chromosomes were able to bind active centromere proteins and provide stable chromosome inheritance for many cell divisions.

(v) *Synthetic centromeres*

The production and successful introduction of the first-generation human artificial chromosome vectors into human cells were the beginnings of synthetic biology methods in centromere research. Synthetic arrays of alpha satellite DNA were generated in part to improve the efficiency of centromere seeding and partly to obviate the need to experimentally manipulate large tracts of satellite repeat DNA (Harrington et al. 1997; Basu et al. 2005). However, questions remain whether these artificial chromosomes are structurally stable in long-term cell culture (Rudd et al. 2003a).

Another approach used in early studies was to retrofit alpha satellite-containing YACs with telomeres and an antibiotic resistance gene (Ikeno et al. 1998). Transfected alpha satellite-retrofitted YAC clones were able to attract key centromere proteins and confer chromosome segregation activity. These transfection experiments showed that centromere DNA sequences, in particular those containing the 17-bp CENP-B box motif that binds directly to the CENP-B protein (Masumoto et al. 1989), were sufficient to attract functional centromere proteins. CENP-B is a transposon-derived gene, related to the Tigger/Pogo family (Smit and Riggs 1996; Kipling and Warburton 1997). CENP-B gene knockout experiments in mouse have shown that it is not required for cell viability or development (Hudson et al. 1998; Kapoor et al. 1998; Perez-Castro et al. 1998) though, paradoxically, it plays an important part in the *de novo* formation of active centromeres in humans and mice and suppresses the formation of extra centromeres on chromosomes (Ohzeki et al. 2002; Okada et al. 2007).

(vi) *Ectopic expression of centromere determinants*

Unlike the canonical histone H3 in plants and animals, CenH3 is loaded onto centromeres independently of DNA replication (Malik and Henikoff 2009). When CenH3 is overexpressed, it creates a condition for it to be mis-loaded into non-centromeric regions, as has been observed in human and *Drosophila* cell lines (Van Hooser et al. 2001; Heun et al. 2006). The mis-targeted

CenH3 is able to attract a subset of kinetochore proteins and, in the face of co-existing native centromeres, triggers chromosome missegregation abnormalities.

In order to alleviate some of the chromosome missegregation defects seen in the continuous overexpression experiments, Olszak and co-workers monitored the short-term overexpression of *Drosophila* CenH3 and the seeding of new centromeres (Olszak et al. 2011). Ectopic centromeres preferentially seeded at the boundaries of euchromatin and heterochromatin, around heterochromatic loci such as those of the pericentric and telomeric regions. These ectopic centromeres are able to attract downstream kinetochore proteins and cause chromosome segregation defects due to dicentric activity. However, the long-term stability of the newly formed kinetochores was not assessed.

(vii) Tethering of CENPs to non-centromeric regions

In other synthetic biology approaches centromere proteins have been redirected to non-centromeric sites by fusing centromere proteins to the lac repressor protein, LacI. These non-centromeric sites are engineered to contain arrays of the lacO DNA repeats which bind to the CENP-LacI fusion protein. In *Drosophila* S2 cells, Mendiburo and colleagues were able to form functional kinetochores by tethering the CenH3-GFP-LacI fusion protein to the lacO array (Mendiburo et al. 2011). Episomal plasmids were able to segregate with moderate stability even after the removal of the CenH3-GFP-LacI fusion protein. Similarly, the CENP-A chaperone, HJURP, was tethered to a non-centromeric site in mammalian cells showing that CENP-A and downstream centromere proteins were able to seed a functional kinetochore (Barnhart et al. 2011). In another approach in chicken and human cells, Gascoigne and colleagues bypassed CENP-A altogether, by tethering other centromere DNA-binding proteins, CENP-C and CENP-T (Gascoigne et al. 2011). The DNA-binding domains were replaced with the lac repressor protein, LacI, which could then bind to an array of lacO repeats at a non-centromeric chromosomal region. This forced localisation of CENP-

C and CENP-T was sufficient to attract downstream kinetochore proteins and partially induce centromere function.

From the above examples of strategies that have been used successfully for the birthing of new centromeres *in vivo* and in the laboratory, it is clear that new centromeres can be formed adaptively from a source of DNA that is related to the natural centromere or not all related to it, or from various ectopically expressed centromere-related chromatin proteins. We remain unclear about the exact mechanisms that trigger the birth of a new centromere but significant strides have been made in the understanding of the essential components and roles of the centromere DNA and its many associated chromatin proteins.

Growth period

Evolutionary New Centromeres (ENCs)

Identification of ENCs by karyotype configuration

Changes in chromosome numbers and configurations are hallmarks that separate species (Coghlan et al. 2005). Some of these changes involve intra and inter chromosome breakage and joining, which in turn changes the configuration of the chromosome with respect to the centromere. A surprising feature of chromosome evolution was the discovery of centromere repositioning along a single chromosome. Evidence to support this idea first emerged when bacterial artificial chromosome (BAC) FISH probes were used to map marker positions along chromosome 9 homologues in primate species (Montefalcone et al. 1999). Further molecular cytogenetic studies have revealed additional centromere position shift events in other primate chromosomes, and in the chromosomes of many other species (Rocchi et al. 2012) (see Birth section). Unlike neocentromeres described in human patients, ENCs have undergone thousands to millions of years of growth and maturation. Since the first report, several questions have remained unanswered, such as: a) whether the centromere repositioning events truly represented the birth of a new centromere or the

transposition of part of the old centromere locus to a new position, and b) did these new centromeres contain canonical centromere repeats, or were they centromere DNA repeat free? Suggestive cues had appeared in the early literature, such as the identification of an anaphoid centromere on the polymorphic chromosome 12 in the orang-utan (Miller et al. 1988). But it wasn't until the genome-sequencing era that the molecular characteristics of ENC's could be verified by CENP-A chromatin immunoprecipitation (ChIP) mapping onto a sequenced and assembled genome (see below).

Satellite-free ENC's

One of the big surprises that have arisen out of the mapping of ENC's by CENP-A ChIP was the discovery of satellite DNA-free centromeres in extant species. To date, such satellite DNA-free centromeres have been reported in orang-utan, horse and chicken (Wade et al. 2009; Shang et al. 2010; Locke et al. 2011).

The orang-utan karyotype contains a polymorphic chromosome 12 with a putative pericentric inversion in both *Pongo abelii* (Sumatran) and *Pongo pygmaeus* (Bornean) species (Dutrillaux et al. 1975; Seuanez et al. 1976). These species diverged around 0.4 to one million years ago (MYA). BAC FISH mapping on the two chromosome 12s showed that there was no change in the chromosomal position of the BAC probes (Locke et al. 2011). This result suggested no evidence for a pericentric inversion, raising the possibility that the observed centromere repositioning was likely due to the appearance of an ENC. CENP-A ChIP mapping indeed confirmed this possibility and the fact that the ENC did not contain any centromere satellite DNA. It is thus somewhat remarkable that the ENC has not acquired any centromere satellite DNA for at least 0.4 million years and possibly up to 12 – 16 million years since the divergence of orang-utans from other primates. Interestingly, the ENC chromosome has at the same time completely lost its old satellite-repeat centromere, with no trace of satellite DNA array detectable by FISH. The acquisition of the

neocentromeric ENC could be explained by an unequal crossing over event, which had deleted the chromosome-12 alpha satellite array.

The horse karyotype also contains a repeat DNA-free ENC, which is not polymorphic, on chromosome 11 (Wade et al. 2009). Mapping of this centromere shows that it contains two CENP-A domains with no other remarkable sequence features. The genus, *Equus* consists of the horse, donkey and zebra, which last shared a common ancestor around three MYA. The donkey and zebra last shared a common ancestor around one MYA. It is interesting to note that the donkey has six centromere repositioning events (Carbone et al. 2006). The horse chromosome-11 ENC is present in horse but not donkey or zebra, dating its appearance during the last three million years.

Chicken centromere DNA has recently been identified by CENP-A ChIP cloning and sequencing (Shang et al. 2010). This DNA is made up of a mixture of satellite DNA and retrotransposons, a feature that finds parallels in other species such as rice (see Maturity/Perpetuity section). CENP-A ChIP sequencing has uncovered an unexpected finding that three centromeres, at chromosomes 5, 27 and Z, were free of any tandemly repeated DNA; however, the sequences of chromosome 27 and Z centromeres did reveal an enrichment for retrotransposons. Furthermore, the CENP-A domain size was found to be atypically small, at around 30 kb. These atypically small kinetochores were functional though not entirely surprising given that chicken kinetochores only bind approximately four spindle microtubules (Ribeiro et al. 2009).

The transition from repeat-free to repeat-occupied centromeres

Once a centromere has shifted to a new position, it has the opportunity or necessity to conform to the genetic and epigenetic characteristics of the established centromeres. It appears that the general trend is for most ENCs that have been identified in mammals to acquire centromeric satellite repeat DNA (Rocchi et al. 2012). A reported case of a centromere that appears to be in transition from a repeat-free to repeat-occupied status is found in the rice chromosome-8 centromere. This

centromere exemplifies a neocentromere that is in the process of accumulating centromeric satellite and centromere-specific retrotransposon DNA. This centromere is thought to have arisen sometime during the last five million years (Nagaki et al. 2004; Yan et al. 2008). CenH3 ChIP sequencing has revealed a 70-kb segment of centromeric satellite DNA embedded within a centromere domain that spans a gene-poor region of 750 kb (Yan et al. 2008). In addition to the centromeric satellite DNA, an ancestral satellite DNA array is found located a few megabases away. Whether these neighbouring array sequences have directly contributed to the formation of the neocentromere and its subsequent acquisition of centromeric repeats is unknown.

The mechanisms governing the transition from repeat-free centromeres to fully mature repeat-rich centromeres is not well understood but are likely to involve a transpositional seeding of sequence materials from a mature centromere to the new centromere via transposable elements, recombination, or rolling circle replication and reinsertion (Plohl et al. 2008). Once the ENC has been seeded, it is then able to expand through additional mechanisms of DNA recombination (see Maturity/Perpetuity below).

An example of centromere repositioning that illustrates the gain and loss of centromere repeat sequences has been described in two closely related plant species, the cucumber and melon (Han et al. 2009), which last shared a common ancestor around nine MYA. In this example, the newly repositioned centromere has acquired the canonical centromere repeat DNA whilst the inactivated centromere has lost its centromere repeats and pericentric heterochromatin (see Death below).

Maturity/Perpetuity

Repeat-rich regional centromeres

Centromere DNA sequences were one of the first genomic regions of multicellular organisms to be identified and characterised. This was mainly due to their abundance, simple tandem repeat organisation, and periodicity of common restriction enzyme sites (Southern 1975). This relatively

uncomplicated organisation turned out to be a recurring feature found in most regional centromeres of single and multicellular eukaryotes.

In humans, the centromere DNA is made up of head-to-tail tandem arrays of a 171-bp monomer repeat known as alpha satellite that often span several megabases in length (Choo 1997; Rudd et al. 2003b). Centromere DNA array lengths are highly variable, both between homologous and heterologous centromeres. Other mammalian and many eukaryotic species also have a similar tandem-repeat organisation at the centromeres, where the DNA sequences are rapidly evolving between species and within chromosomes of the same species (Plohl et al. 2008).

Some species such as chickens *Drosophila*, and many plants, contain a mixture of tandemly repeated DNA and retrotransposable elements. What is common for all regional centromeres is that the repetitive DNA provides a preferred chromatin environment for the maintenance and perpetuation of the centromeric locus. Key characteristics of this chromatin environment include the presence of heterochromatin, suppression of transcription and meiotic recombination, and enriched sister centromere cohesion as a last point of chromosomal attachment before the onset of anaphase.

Centromere DNA evolution

The rapid evolution of centromere DNA within single centromeres of individual chromosomes and between species presents a remarkable feature of this locus. Most of the proposed molecular mechanisms are non-Mendelian in nature, centred on DNA repair and replication pathways. Unequal crossing over is suggested to account for most of the expansion and variation of centromere DNA sequences (Smith 1976). Other mechanisms that contribute to centromere evolution are replication slippage, sequence (gene) conversion, rolling circle replication and reinsertion, transposition and retrotransposition (Smith 1976; Dover 1982; Pertile et al. 2009; Shi et al. 2010).

ChIP experiments in rice have revealed rapid evolution of repeats in closely related rice species (Lee et al. 2005). In a relatively short evolutionary space of time, 5 to 10 million years, one centromere satellite DNA repeat has completely replaced by other satellite and retrotransposable DNAs. This rapid rate of evolution is compounded in haploid centromeres such as the primate and mouse chromosome Y centromere (Pertile et al. 2009). The observation that the mouse Y centromere displays a stable molecular structure for hundreds of recent generations suggests that rapid centromere DNA changes may occur sporadically, perhaps in response to environmental cues, followed by periods of structural stability through evolution.

Centromere/kinetochore expansion and stabilisation

The centromere DNA is the underlying foundation upon which the mature kinetochore forms. During the non-mitotic phase of the cell cycle, the kinetochore is epigenetically marked by a group of foundation centromere proteins, also known as the Constitutive Centromere Associated Network (CCAN) (Amor et al. 2004; Hori et al. 2008; Santaguida and Musacchio 2009). The repetitive DNA makeup of the centromere is in turn reflected by multiple kinetochore protein complexes, which go on to bind multiple spindle microtubules. Since the kinetochore structure of regional centromeres is made up of multiple subunits of the base centromere protein components then it would be interesting to speculate that the underlying repetitive DNA assists in maintaining this structure between cell divisions.

The structural relationship between the foundation kinetochore protein CenH3 and the centromere DNA has been examined using ChIP and immuno-FISH methods. Analysis in plants (Zhong et al. 2002; Nagaki et al. 2003) and metazoans (Zinkowski et al. 1991; Haaf and Ward 1994; Lo et al. 2001; Blower et al. 2002) indicate that the CenH3 protein occupies only a (relatively small) portion of the centromere chromatin DNA, often showing a discontinuous distribution pattern. It is hypothesized that this discontinuous arrangement of CenH3 assists in the spatial formation of the kinetochore (Zinkowski et al. 1991; Blower et al. 2002; Marshall et al. 2008b). The centromere

DNA flanking the CenH3-occupied domains can be of the same repeat type that is bound by CenH3 or an unrelated repeat DNA that is predominantly found at pericentromeric regions (Rudd et al. 2003b; Malik and Henikoff 2009).

The structural relationship between centromere DNA and the kinetochore is further exemplified in the large compound centromeres in the large chromosomes of the Indian muntjac deer (Vafa et al. 1999). Interestingly, in these chromosomes, centromere DNA expansion appears to have been accompanied by a corresponding expansion of the kinetochore region. Whilst it has been suggested that large kinetochores can presumably arise from multiple rounds of centric fusion during the karyotype evolution of the species (Brinkley et al. 1984), such a mechanism is unlikely to be universal and would not adequately explain situations in which kinetochores have spread out to cover the entire (spindle microtubule binding surface) length of whole chromosomes (Guerra et al. 2010) (see Runaway Centromeres below).

Limiting the centromere from spreading across the whole chromosome

If centromere repeats can expand and contract by mechanisms such as unequal crossing over (Smith 1976), then what intrinsic cellular mechanisms keep the centromere DNA from expanding uncontrollably? There appears to be at least two mechanisms that operate to check any rampant spreading of CenH3 and its overlaying kinetochore into other parts of the genome. These mechanisms relate to the role of pericentromeric heterochromatin and negative selection against mislocalisation of CenH3.

As pericentromere DNA flanking the CenH3 region is generally marked by the constitutive heterochromatic histone modification mark, histone H3 lysine 9 tri-methyl (H3K9me3), together with other heterochromatin-associated proteins. Pericentric heterochromatin has multiple functions, including epigenetic silencing of gene transcription, sister centromere cohesion and, importantly, acting as a barrier between the centromere and the euchromatin. In fission yeast, candidate barrier

sequences in the form of tRNA genes flank the core centromere domains. Deletion of these genes results in the spreading of the flanking pericentric heterochromatin into the CenH3 domain thus weakening centromere function (Scott et al. 2006). Contrastingly, H3K9me3 heterochromatin is needed for the retention of normal amounts of CenH3 at the centromere of the filamentous fungus, *Neurospora crassa*. This heterochromatic mark is present throughout the CenH3 region whereas the H3K4me2 mark that is commonly found in many metazoan centromeres is not present (Blower et al. 2002; Smith et al. 2011). Loss of this silencing mark in mutants of H3K9me transferase and the heterochromatin protein, HP1, showed that CenH3 levels were reduced at centromeres when compared to wild type (Smith et al. 2011).

In human artificial chromosomes in which alpha satellite DNA is interrupted with plasmid vector and an antibiotic resistance gene, the CenH3 is shown to spread into the intervening vector sequences (Lam et al. 2006). The confinement of CenH3 at the centromeres can also be lifted when the protein is overexpressed in human and *Drosophila* cells (Van Hooser et al. 2001; Tomonaga et al. 2003; Heun et al. 2006). CenH3 overexpression results in mislocalisation of the protein and triggers the partial assembly of downstream kinetochore components at ectopic chromosomal sites that in turn attracts spindle microtubules and leads to an increase in chromosome missegregation (Heun et al. 2006; Au et al. 2008; Amato et al. 2009). Such a phenotype points to a negative selection mechanism against CenH3 spreading. In addition, molecular mechanisms exist to ensure that CenH3 is correctly localised at the centromere. Proteasome-mediated degradation of mis-incorporated CenH3 is found in organisms such as budding yeast and flies (Collins et al. 2004; Moreno-Moreno et al. 2006). In yeast, mis-incorporated CenH3 is targeted for proteolysis by the E3 ubiquitin ligase, Psh1 (Ranjitkar et al. 2010).

During the last decade sequence analysis studies on DNA-binding centromere proteins, such as CenH3 and CENP-C, have revealed a paradoxically high rate of evolution for such essential cell cycle proteins (Malik and Henikoff 2001; Talbert et al. 2002; Cooper and Henikoff 2004; Talbert et al. 2004; Schueler et al. 2010). This high rate of sequence divergence mirrors the high rate of

evolutionary change in centromere DNA. Henikoff and colleagues have proposed the “centromere drive” model which describes a mechanism where centromeres are under competitive selection to be included into the oocyte, which undergoes asymmetric meiotic divisions during which three out of four meiotic products are eliminated (Henikoff et al. 2001). Close inspection of the type of sequence changes that are driving centromere protein evolution have shown that they are undergoing adaptive or positive selection (Malik 2009). This is especially evident in organisms with a combination of asymmetric meiotic divisions in females and symmetric meiotic divisions in males. In these organisms, centromeres have been proposed to be under “conflict” where expanding centromere DNA is kept under check by adapting centromere and heterochromatin proteins. In organisms where asymmetric meiotic divisions are absent such as yeast, we do not observe any positive selection for CenH3, and centromeres are kept to a minimum size, since there is no selective advantage for a larger centromere (Talbert et al. 2004; Baker and Rogers 2006). This may also explain why the chromosome Y centromere in humans and mice is smaller than the centromeres of the autosomes and X chromosome, since the Y centromere is only present in meiotic tissues that undergo symmetric cell divisions (Irvine et al. 2004; Pertile et al. 2009). Therefore, one would expect runaway centromere DNA expansions in organisms that have only asymmetric “female” meiotic cell divisions. Recently, CenH3 sequences have been analysed in multiple species of the ciliated protozoan lineage, Tetrahymena (Elde et al. 2011). This single-celled organism only has asymmetric “female” meiotic cell divisions. Surprisingly, it was found that the CenH3 proteins do not show any evidence for positive selection which suggests unsuppressed centromere-drive in the absence of male meiosis. However, no studies have been performed to define the centromere DNA along the chromosomes of these species.

Are non-localised holocentric centromeres runaway centromere expansions?

As mentioned above, holocentric chromosomes contain a kinetochore that spans the entire length of each chromosome. These atypical centromeres have been identified in organisms such as worm,

insects, arachnids, plants and ciliates (Guerra et al. 2010). This type of centromere is not unique in eukaryotes and appears to have arisen multiple times during evolution. So are holocentric centromeres examples of a runaway centromere expansion? Malik and Henikoff propose that holocentricity may have been a mechanism to suppress centromere drive by limiting the spread or accumulation of centromere repeat DNA (Malik 2009). This hypothesis is circumstantially supported by a lack of evidence for any obvious abundant centromere repeat DNA. That said, even though the CenH3 protein has been identified from several holocentric species, the actual centromere DNA remains to be identified. Possibly, the kinetochore binding sequence has completely lost its affinity to satellite or transposon DNA and is primarily directed by epigenetic mechanisms.

Death

Similar to the birth of centromeres, death is often triggered by a chromosomal rearrangement event where, for instance, two centromeres are brought together onto the same chromosome. This can occur via the translocation of two active centromeres to one chromosome, or by the appearance of a neocentromere onto an existing chromosome (Marshall et al. 2008a). In most eukaryotes, when two active centromeres separated by a substantial stretch of intervening DNA can result in the tearing apart of the chromosomes if both centromeres from the same chromatid are not orientated to the same pole (McClintock 1938; Koshland et al. 1987; Lukaszewski 1995). This situation is alleviated in holocentric chromosomes where scaffold proteins such as the condensin complex or other centromere proteins keep the sister chromatids rigid and parallel thus minimising the twisting and incorrect orientation that would result in chromosome breakage (Hagstrom et al. 2002).

To alleviate the breaking of a chromosome with two active centromeres, cells have a mechanism to inactivate one of the centromeres. The exact process governing the choice of which centromere becomes inactive is largely unknown. However, studies with engineered human dicentric

chromosomes have shown that such a process is mostly epigenetic and that the inactivation status is not necessarily fixed (Higgins et al. 2005).

Once a centromere is no longer functional, what then happens to its centromere proteins? In humans, functionally essential centromere proteins no longer bind to inactivated centromeres, with the exception of CENP-B and some limited binding of heterochromatin proteins (Earnshaw and Migeon 1985; Page et al. 1995; Sullivan and Schwartz 1995). The consequence of the loss of active centromere marks is the abolition of spindle microtubule binding and a decrease in sister centromere cohesion late in metaphase. In the maize plant, centromere inactivation of dicentric chromosomes also involves the loss of the CenH3 mark (Han et al. 2006).

Once a centromere has “died”, there will no longer be any requirement to maintain its centromere DNA sequences. With time, these sequences will decay away due to deletions, the accumulation of mutations, and the insertion of transposable elements. During the evolution of humans, there have been several instances where centromeres have undergone inactivation. Known chromosomal sites showing remnants of past centromere inactivation that have been reported include 2q21, 6p22, 9q13 and 15q25 (Yunis and Prakash 1982; Baldini et al. 1993; Ventura et al. 2003; Capozzi et al. 2009). For example, after the divergence of the hominid and chimp lineages, two chromosomes have undergone a fusion translocation to produce human chromosome 2 (Yunis and Prakash 1982). FISH mapping studies with centromeric alpha satellite DNA showed the existence of a secondary hybridisation signal on the long arm of chromosome 2 (Baldini et al. 1993). This was the remnant of a previously active centromere, which is now decaying away. Since the human genome has almost been fully sequenced and assembled (apart from heterochromatin and centromere regions which are composed of homogeneous arrays of satellite repeats), we closely examined this previously active region. This region showed a stretch of alpha satellite spanning 42 kb at chr2:132966375-133007983 (GRCh37/hg19 assembly). Dot plot alignment analysis showed no evidence of any higher-order repeat structure. Our analysis has also detected the presence of recently inserted interspersed repeats including a full length LINE 1 (class L1PA3) and a hominid-

specific retroposon, SVA. This once-active centromere is present in other primate species such as chimp, gorilla and orang-utan. Based on the sequenced chimp genome and partially assembled centromere-flanking regions of chromosome 2b chr2B:132,834,456-132,900,124 and chr2B:135,910,915- 136,163,394, p and q side respectively (CGSC 2.1.3/panTro3 assembly), there is some evidence for a higher-order repeat structure, with a 1.2 kb and 0.34 – 0.68 kb repeating unit, at the p and q flanking regions, respectively. These analyses indicate that the death of a centromere is accompanied by a significant loss of centromere DNA composition and organisation.

Resurrection

The two most common types of DNA repeats to be found at centromeres are satellite and retrotransposon DNAs. These sequences are not necessarily restricted to centromeric regions, as has been shown in multiple organisms (Sun et al. 1997; Rudd and Willard 2004; Ma et al. 2007; Cellamare et al. 2009; Nagaki et al. 2009; Shang et al. 2010). This observation begs the question of what is the possible role of such sequences away from their centromeric home. Are they simply vestiges of old, inactive centromeres, mislocated centromeres or have they been borrowed from other regions such as telomeres as observed in *Drosophila* (Sun et al. 1997)?

Another possible role for these sequences is in the seeding or “resurrection” of new centromeres. One example where an inactivated centromere can be resurrected has been described in maize (Han et al. 2009). In this study, unstable dicentric chromosomes with a large and small centromere with typical centromere DNAs have become stabilised through the inactivation of the smaller centromere. The small centromere stably retains its inactive state through multiple generations. To test whether the inactive small centromere could be reactivated, the investigators induced an intrachromosomal break to separate the two centromeres. Once separated the previously inactivated centromere became active, as shown by the functional centromere marks, CenH3 and CENP-C, and the stable segregation of its associated chromosome. What remains unanswered is whether the inactivated centromere contains any unidentified epigenetic marks that keep the centromere in a

dormant state, or whether the relatively intact DNA sequence and topology are sufficient to trigger reactivation (see also Birth of centromere via naked centromere DNA).

ENCs or neocentromeres with flanking centromere DNA?

Presently, there is no clear evidence to suggest that ENCs have arisen from inactivated centromere DNA. It is difficult to track the life history on an ENC because its neocentric event has usually occurred thousands to million of years ago, allowing new centromeric DNA ample opportunity to invade the ENC site (see Growth above). Even the reporting of many *de novo* neocentromere cases in humans has provided no obvious “resurrection” event from previously inactivated alphoid DNA (Marshall et al. 2008a). One possible primate group that lends itself for the investigation of the re-birth of new centromeres is the lesser apes. This group of species is known to have undergone rapid karyotype evolution (Misceo et al. 2008). One striking example is in the white-cheeked gibbon, *Nomascus leucogenys*. Cloning and FISH analysis of the centromeric alpha satellite has shown a complex genomic distribution, where strong alphoid signals are not only found at centromeres, but telomeres and interstitial regions (Cellamare et al. 2009). This may reflect genomic rearrangements across alphoid regions and the resulting greater opportunity for the reactivation of the inactivated alphoid regions.

Clustering of human clinical neocentromeres on the q-arm of chromosome 15 has suggested the possibility of a sequence dependent seeding site. Closer inspection of this region reveals the presence of an ancestral centromere containing segmentally duplicated DNA clusters that typically reside around pericentric regions (Ventura et al. 2003; She et al. 2004). This ancient centromere was estimated to be inactivated around 25 MYA, after the divergence of the great apes from old world monkeys (Ventura et al. 2003). Chromosome 6p22.1 is another possible site of centromere rebirth because of the presence of an ancient centromere, however only one case has been described (Capozzi et al. 2009). Evidence to suggest that inactive alpha satellite may not form neocentromeres comes from the largest non-centromeric block of alpha satellite on 2q21.1 (see Death section)

where no neocentromere cases have been reported, though this could be due to the relatively low number of clinical neocentromere cases described to date, or the presence of important genes whose triple dosage is incompatible with normal development.

Conclusions

The resilient centromere

The centromere is an intriguing chromosomal structure that is capable of undergoing evolutionary cycles of birth, growth, maturation, death and resurrection. Such a “centromere life cycle” characteristic is not entirely surprising given the biologically important roles centromeres play in protecting the proper inheritance of our DNA, and the necessity to maintain its perpetual presence in the face of untold genomic rearrangements that drive evolution on the one hand and threatens the ongoing existence of the centromere on the other hand. Paradoxically, for such an essential functional role, the centromere evolves rapidly at the DNA and protein level. This high rate of change attests to the remarkable resilience of the centromere as it constantly adapts to its genomic environment. A key contributor to this resilience rests in the epigenetic nature of the centromere. It is this epigenetic capability that allows an inactivated centromere or the chromatin of an ancient or hitherto non-centromeric genomic site to be remodelled into a functional kinetochore. However, we are far from fully grasping the nature of the epigenetic mechanisms at play. The roles of DNA methylation, histone modification, centromere transcription and its transcripts, are only beginning to be recognised. The detailed contributions of these players remain to be deciphered and present fertile grounds for research for the cellular, molecular and evolutionary biologists.

Acknowledgements

This work was supported by the National Health and Medical Research Council (NHMRC) of Australia and the Victorian Government’s Operational Infrastructure Support Programme. PK was

supported by an R.D. Wright Fellowship, and KHAC by a Senior Principal Research Fellowship of NHMRC.

Figure

Fig. 1. Diagram showing various aspects of the centromere life cycle. Birth: chromosome containing a mature, repeat-rich centromere (green box) undergoes fragmentation, followed by the epigenetic assembly of a neocentromere (yellow box). Growth: neocentromere gradually accumulates some centromere repeat DNA, or remains centromere repeat-free. Maturity/Perpetuity: centromere is saturated with repeat sequences and undergoes rounds of contractions and expansions, together with rapid evolution of the underlying repeats. Death: genomic rearrangements can produce neocentromere-centromere or centromere-centromere functional dicentric chromosomes. One of the active centromeres is switched off and dies (black box). Resurrection: chromosome with two repeat-rich centromeres (one active and one inactive) is fragmented, followed by epigenetic resurrection of the dead centromere to provide mitotic stability to the otherwise acentric chromosomal fragment.

Table

Table 1: Assisted centromere production

Type	Method	Organism	Centromere activity (chromosome stability)	Reference
Removal or disruption of endogenous centromere	Radiation DNA break	<i>Drosophila</i>	Very high	(Williams et al. 1998)
	DNA recombination	<i>Drosophila</i>	Very high	(Platero et al. 1999)
		<i>S. pombe</i>	Very high	(Ishii et al. 2008)
	Chromosomal rearrangements	Barley	Very high	(Nasuda et al. 2005)
	Centromere DNA insertion	<i>Candida albicans</i>	Very high	(Nasuda et al. 2005)
Centromere DNA transformation/transfection	Plasmid	<i>S. cerevisiae</i>	Very high	(Fitzgerald-Hayes et al. 1982)
		<i>Kluyveromyces lactis</i>	Very high	(Heus et al. 1990)
		<i>Candida maltosa</i>	Very high	(Ohkuma et al. 1995)
		<i>Candida glabrata</i>		
		<i>S. pombe</i>		(Stoyan and

				Carbon 2004) (Hahnenberger et al. 1989)
	Linear chromosome	<i>S. cerevisiae</i> <i>S. pombe</i>	Very high Very high	(Dani and Zakian 1983) (Hahnenberger et al. 1989)
	Insertion	Human		(Haaf et al. 1992; Larin et al. 1994)
Synthetic centromeres/chromosomes	Tandem repeat concatenation and transfection	Human	High	(Harrington et al. 1997)
	YAC retrofit	Human	High	(Ikeno et al. 1998)
	Tandem repeat concatenation and ligation into BAC vector Circular/linear (Basu)	Human	High	(Ohzeki et al. 2002; Basu et al. 2005; Okada et al. 2007)
Ectopic/overexpression	CENP-A conditional overexpression with CMV-based	Human	Partial	(Van Hooser et al. 2001)

	tetracycline regulated promoter			
	Cid conditional or ubiquitous overexpression	<i>Drosophila</i>	Partial	(Heun et al. 2006)
	Cid transient expression	<i>Drosophila</i>	Medium	(Olszak et al. 2011)
Centromere protein tethering	Cid-GFP-LacI mCherry-LacI- HJURP GFP-CENP-C- Δ C-LacI and – GFP-CENP-T- Δ C-LacI tethering to DNA without CENP-A	<i>Drosophila</i> Mammalian Human and chicken	Medium Partial Partial	(Mendiburo et al. 2011) (Barnhart et al. 2011) (Gascoigne et al. 2011)

References

- Allshire RC, Karpen GH (2008) Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat Rev Genet* 9 (12):923-937
- Amato A, Schillaci T, Lentini L, Di Leonardo A (2009) CENPA overexpression promotes genome instability in pRb-depleted human cells. *Mol Cancer* 8:119
- Amor DJ, Kalitsis P, Sumer H, Choo KHA (2004) Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol* 14 (7):359-368
- Au WC, Crisp MJ, DeLuca SZ, Rando OJ, Basrai MA (2008) Altered dosage and mislocalization of histone H3 and Cse4p lead to chromosome loss in *Saccharomyces cerevisiae*. *Genetics* 179 (1):263-275
- Baker RE, Rogers K (2006) Phylogenetic analysis of fungal centromere H3 proteins. *Genetics* 174 (3):1481-1492
- Baldini A, Ried T, Shridhar V, Ogura K, D'Aiuto L, Rocchi M, Ward DC (1993) An alphoid DNA sequence conserved in all human and great ape chromosomes: evidence for ancient centromeric sequences at human chromosomal regions 2q21 and 9q13. *Hum Genet* 90 (6):577-583
- Barnhart MC, Kuich PH, Stellfox ME, Ward JA, Bassett EA, Black BE, Foltz DR (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol* 194 (2):229-243
- Basu J, Stromberg G, Compitello G, Willard HF, Van Bokkelen G (2005) Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. *Nucleic Acids Res* 33 (2):587-596
- Baum M, Sanyal K, Mishra PK, Thaler N, Carbon J (2006) Formation of functional centromeric chromatin is specified epigenetically in *Candida albicans*. *Proc Natl Acad Sci U S A* 103 (40):14877-14882

- Black BE, Cleveland DW (2011) Epigenetic centromere propagation and the nature of CENP-a nucleosomes. *Cell* 144 (4):471-479
- Blower MD, Sullivan BA, Karpen GH (2002) Conserved organization of centromeric chromatin in flies and humans. *Dev Cell* 2 (3):319-330
- Brinkley BR, Valdivia MM, Tousson A, Brenner SL (1984) Compound kinetochores of the Indian muntjac. Evolution by linear fusion of unit kinetochores. *Chromosoma* 91 (1):1-11
- Capozzi O, Purgato S, D'Addabbo P, Archidiacono N, Battaglia P, Baroncini A, Capucci A, Stanyon R, Della Valle G, Rocchi M (2009) Evolutionary descent of a human chromosome 6 neocentromere: a jump back to 17 million years ago. *Genome Res* 19 (5):778-784
- 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, Guerin G, Archidiacono N, Rocchi M, Giulotto E (2006) Evolutionary movement of centromeres in horse, donkey, and zebra. *Genomics* 87 (6):777-782
- Cellamare A, Catacchio CR, Alkan C, Giannuzzi G, Antonacci F, Cardone MF, Della Valle G, Malig M, Rocchi M, Eichler EE, Ventura M (2009) New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. *Mol Biol Evol* 26 (8):1889-1900
- Choo KHA (1997) *The Centromere*. Oxford University Press, Oxford, New York, Tokyo
- Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet* 21 (12):673-682
- Collins KA, Furuyama S, Biggins S (2004) Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. *Curr Biol* 14 (21):1968-1972
- Cooper JL, Henikoff S (2004) Adaptive evolution of the histone fold domain in centromeric histones. *Mol Biol Evol* 21 (9):1712-1718
- Dani GM, Zakian VA (1983) Mitotic and meiotic stability of linear plasmids in yeast. *Proc Natl Acad Sci U S A* 80 (11):3406-3410

- Dover G (1982) Molecular drive: a cohesive mode of species evolution. *Nature* 299 (5879):111-117
- Dutrillaux B, Rethore MO, Lejeune J (1975) [Comparison of the karyotype of the orangutan (*Pongo pygmaeus*) to those of man, chimpanzee, and gorilla]. *Ann Genet* 18 (3):153-161
- Earnshaw WC, Migeon BR (1985) Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma* 92 (4):290-296
- Elde NC, Roach KC, Yao MC, Malik HS (2011) Absence of positive selection on centromeric histones in *Tetrahymena* suggests unsuppressed centromere: drive in lineages lacking male meiosis. *J Mol Evol* 72 (5-6):510-520
- Fitzgerald-Hayes M, Clarke L, Carbon J (1982) Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. *Cell* 29 (1):235-244
- Gascoigne KE, Takeuchi K, Suzuki A, Hori T, Fukagawa T, Cheeseman IM (2011) Induced ectopic kinetochore assembly bypasses the requirement for CENP-A nucleosomes. *Cell* 145 (3):410-422
- Guerra M, Cabral G, Cuacos M, Gonzalez-Garcia M, Gonzalez-Sanchez M, Vega J, Puertas MJ (2010) Neocentrics and holokinetics (holocentrics): chromosomes out of the centromeric rules. *Cytogenet Genome Res* 129 (1-3):82-96
- Haaf T, Warburton PE, Willard HF (1992) Integration of human alpha-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal chromosome segregation. *Cell* 70 (4):681-696
- Haaf T, Ward DC (1994) Structural analysis of alpha-satellite DNA and centromere proteins using extended chromatin and chromosomes. *Hum Mol Genet* 3 (5):697-709
- Hagstrom KA, Holmes VF, Cozzarelli NR, Meyer BJ (2002) *C. elegans* condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev* 16 (6):729-742

- Hahnenberger KM, Baum MP, Polizzi CM, Carbon J, Clarke L (1989) Construction of functional artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. *Proc Natl Acad Sci U S A* 86 (2):577-581
- Han F, Lamb JC, Birchler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc Natl Acad Sci U S A* 103 (9):3238-3243
- Han Y, Zhang Z, Liu C, Liu J, Huang S, Jiang J, Jin W (2009) Centromere repositioning in cucurbit species: implication of the genomic impact from centromere activation and inactivation. *Proc Natl Acad Sci U S A* 106 (35):14937-14941
- Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet* 15 (4):345-355
- Hassold T, Hunt P (2001) To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* 2 (4):280-291
- Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293 (5532):1098-1102.
- Heun P, Erhardt S, Blower MD, Weiss S, Skora AD, Karpen GH (2006) Mislocalization of the *Drosophila* centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell* 10 (3):303-315
- Heus JJ, Zonneveld BJ, de Steensma HY, van den Berg JA (1993) The consensus sequence of *Kluyveromyces lactis* centromeres shows homology to functional centromeric DNA from *Saccharomyces cerevisiae*. *Mol Gen Genet* 236 (2-3):355-362
- Heus JJ, Zonneveld BJ, Steensma HY, Van den Berg JA (1990) Centromeric DNA of *Kluyveromyces lactis*. *Curr Genet* 18 (6):517-522
- Higgins AW, Gustashaw KM, Willard HF (2005) Engineered human dicentric chromosomes show centromere plasticity. *Chromosome Res* 13 (8):745-762

- Hori T, Amano M, Suzuki A, Backer CB, Welburn JP, Dong Y, McEwen BF, Shang WH, Suzuki E, Okawa K, Cheeseman IM, Fukagawa T (2008) CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell* 135 (6):1039-1052
- Hudson DF, Fowler KJ, Earle E, Saffery R, Kalitsis P, Trowell H, Hill J, Wreford NG, de Kretser DM, Cancilla MR, Howman E, Hii L, Cutts SM, Irvine DV, Choo KHA (1998) Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. *J Cell Biol* 141 (2):309-319
- Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat Biotechnol* 16 (5):431-439
- Irvine DV, Amor DJ, Perry J, Sirvent N, Pedoutour F, Choo KHA, Saffery R (2004) Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres. *Chromosome Res* 12 (8):805-815
- Ishii K, Ogiyama Y, Chikashige Y, Soejima S, Masuda F, Kakuma T, Hiraoka Y, Takahashi K (2008) Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. *Science* 321 (5892):1088-1091
- Kapoor M, Montes de Oca Luna R, Liu G, Lozano G, Cummings C, Mancini M, Ouspenski I, Brinkley BR, May GS (1998) The cenpB gene is not essential in mice. *Chromosoma* 107 (8):570-576.
- Ketel C, Wang HS, McClellan M, Bouchonville K, Selmecki A, Lahav T, Gerami-Nejad M, Berman J (2009) Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. *PLoS Genet* 5 (3):e1000400
- Kipling D, Warburton PE (1997) Centromeres, CENP-B and Tigger too. *Trends Genet* 13 (4):141-145

- Kitada K, Yamaguchi E, Arisawa M (1996) Isolation of a *Candida glabrata* centromere and its use in construction of plasmid vectors. *Gene* 175 (1-2):105-108
- Koshland D, Rutledge L, Fitzgerald-Hayes M, Hartwell LH (1987) A genetic analysis of dicentric minichromosomes in *Saccharomyces cerevisiae*. *Cell* 48 (5):801-812
- Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA (2006) Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc Natl Acad Sci U S A* 103 (11):4186-4191
- Larin Z, Fricker MD, Tyler-Smith C (1994) De novo formation of several features of a centromere following introduction of a Y alphoid YAC into mammalian cells. *Hum Mol Genet* 3 (5):689-695
- Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. *Proc Natl Acad Sci U S A* 102 (33):11793-11798
- Lo AWI, Craig JM, Saffery R, Kalitsis P, Irvine DV, Earle E, Magliano DJ, Choo KHA (2001) A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *Embo J* 20 (8):2087-2096.
- Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chinwalla AT, Minx P, Mitreva M, Cook L, Delehaunty KD, Fronick C, Schmidt H, Fulton LA, Fulton RS, Nelson JO, Magrini V, Pohl C, Graves TA, Markovic C, Cree A, Dinh HH, Hume J, Kovar CL, Fowler GR, Lunter G, Meader S, Heger A, Ponting CP, Marques-Bonet T, Alkan C, Chen L, Cheng Z, Kidd JM, Eichler EE, White S, Searle S, Vilella AJ, Chen Y, Flicek P, Ma J, Raney B, Suh B, Burhans R, Herrero J, Haussler D, Faria R, Fernando O, Darre F, Farre D, Gazave E, Oliva M, Navarro A, Roberto R, Capozzi O, Archidiacono N, Della Valle G, Purgato S, Rocchi M, Konkel MK, Walker JA, Ullmer B, Batzer MA, Smit AF, Hubley R, Casola C, Schridder DR, Hahn MW, Quesada V, Puente XS, Ordonez GR, Lopez-Otin C, Vinar T, Brejova B, Ratan A, Harris RS, Miller W, Kosiol C, Lawson HA,

- Taliwal V, Martins AL, Siepel A, Roychoudhury A, Ma X, Degenhardt J, Bustamante CD, Gutenkunst RN, Mailund T, Dutheil JY, Hobolth A, Schierup MH, Ryder OA, Yoshinaga Y, de Jong PJ, Weinstock GM, Rogers J, Mardis ER, Gibbs RA, Wilson RK (2011) Comparative and demographic analysis of orang-utan genomes. *Nature* 469 (7331):529-533
- Lukaszewski AJ (1995) Chromatid and chromosome type breakage-fusion-bridge cycles in wheat (*Triticum aestivum* L.). *Genetics* 140 (3):1069-1085
- Ma J, Wing RA, Bennetzen JL, Jackson SA (2007) Plant centromere organization: a dynamic structure with conserved functions. *Trends Genet* 23 (3):134-139
- Malik HS (2009) The centromere-drive hypothesis: a simple basis for centromere complexity. *Prog Mol Subcell Biol* 48:33-52
- Malik HS, Henikoff S (2001) Adaptive evolution of Cid, a centromere-specific histone in *Drosophila*. *Genetics* 157 (3):1293-1298
- Malik HS, Henikoff S (2009) Major evolutionary transitions in centromere complexity. *Cell* 138 (6):1067-1082
- Marshall OJ, Choo KH (2009) Neocentromeres come of age. *PLoS Genet* 5 (3):e1000370
- Marshall OJ, Chueh AC, Wong LH, Choo KHA (2008a) Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet* 82 (2):261-282
- Marshall OJ, Marshall AT, Choo KHA (2008b) Three-dimensional localization of CENP-A suggests a complex higher order structure of centromeric chromatin. *J Cell Biol* 183 (7):1193-1202
- Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol* 109 (5):1963-1973

- McClintock B (1938) The production of homozygous deficient tissues with mutant characteristics by means of the aberrant mitotic behaviour of ring-shaped chromosomes. *Genetics* 23:315-376
- Mendiburo MJ, Padeken J, Fulop S, Schepers A, Heun P (2011) *Drosophila* CENH3 is sufficient for centromere formation. *Science* 334 (6056):686-690
- Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol* 7 (3):R23
- Miller DA, Sharma V, Mitchell AR (1988) A human-derived probe, p82H, hybridizes to the centromeres of gorilla, chimpanzee, and orangutan. *Chromosoma* 96 (4):270-274
- Misceo D, Capozzi O, Roberto R, Dell'oglio MP, Rocchi M, Stanyon R, Archidiacono N (2008) Tracking the complex flow of chromosome rearrangements from the Hominoidea Ancestor to extant *Hylobates* and *Nomascus* Gibbons by high-resolution synteny mapping. *Genome Res* 18 (9):1530-1537
- Montefalcone G, Tempesta S, Rocchi M, Archidiacono N (1999) Centromere repositioning. *Genome Res* 9 (12):1184-1188
- Moreno-Moreno O, Torras-Llort M, Azorin F (2006) Proteolysis restricts localization of CID, the centromere-specific histone H3 variant of *Drosophila*, to centromeres. *Nucleic Acids Res* 34 (21):6247-6255
- Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* 36 (2):138-145
- Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. *Genetics* 163 (3):1221-1225
- Nagaki K, Walling J, Hirsch C, Jiang J, Murata M (2009) Structure and evolution of plant centromeres. *Prog Mol Subcell Biol* 48:153-179

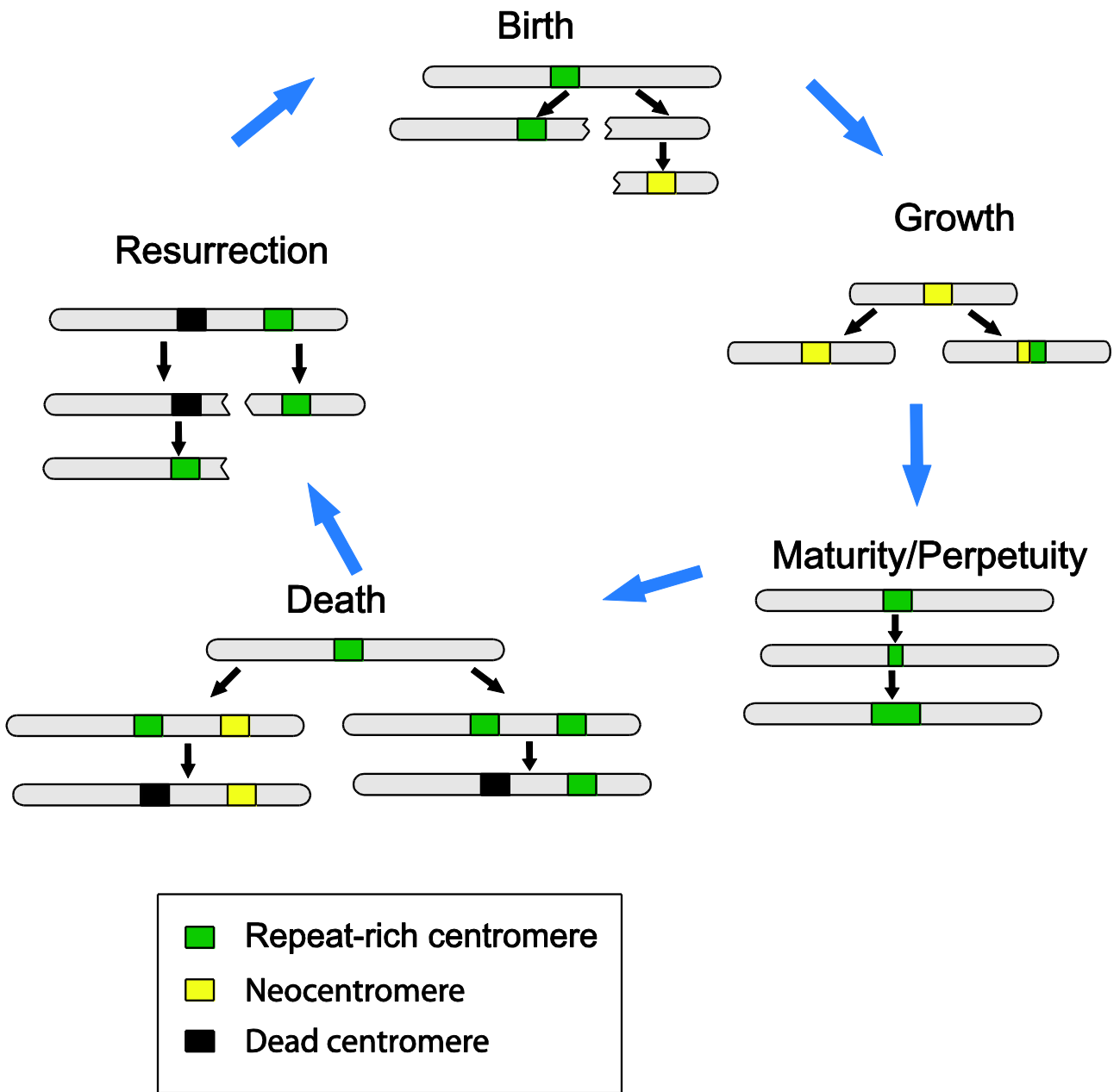
- Nasuda S, Hudakova S, Schubert I, Houben A, Endo TR (2005) Stable barley chromosomes without centromeric repeats. *Proc Natl Acad Sci U S A* 102 (28):9842-9847
- Ohkuma M, Kobayashi K, Kawai S, Hwang CW, Ohta A, Takagi M (1995) Identification of a centromeric activity in the autonomously replicating TRA region allows improvement of the host-vector system for *Candida maltosa*. *Mol Gen Genet* 249 (4):447-455
- Ohzeki J, Nakano M, Okada T, Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol* 159 (5):765-775
- Okada T, Ohzeki J, Nakano M, Yoda K, Brinkley WR, Larionov V, Masumoto H (2007) CENP-B controls centromere formation depending on the chromatin context. *Cell* 131 (7):1287-1300
- Olszak AM, van Essen D, Pereira AJ, Diehl S, Manke T, Maiato H, Saccani S, Heun P (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol* 13 (7):799-808
- Page SL, Earnshaw WC, Choo KHA, Shaffer LG (1995) Further evidence that CENP-C is a necessary component of active centromeres: studies of a dic(X; 15) with simultaneous immunofluorescence and FISH. *Hum Mol Genet* 4 (2):289-294
- Perez-Castro AV, Shamanski FL, Meneses JJ, Lovato TL, Vogel KG, Moyzis RK, Pedersen R (1998) Centromeric protein B null mice are viable with no apparent abnormalities. *Dev Biol* 201 (2):135-143.
- Perpelescu M, Fukagawa T (2011) The ABCs of CENPs. *Chromosoma*
- Pertile MD, Graham AN, Choo KHA, Kalitsis P (2009) Rapid evolution of mouse Y centromere repeat DNA belies recent sequence stability. *Genome Res* 19 (12):2202-2213
- Platero JS, Ahmad K, Henikoff S (1999) A distal heterochromatic block displays centromeric activity when detached from a natural centromere. *Mol Cell* 4 (6):995-1004
- Plohl M, Luchetti A, Mestrovic N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409 (1-2):72-82

- Ranjitkar P, Press MO, Yi X, Baker R, MacCoss MJ, Biggins S (2010) An E3 ubiquitin ligase prevents ectopic localization of the centromeric histone H3 variant via the centromere targeting domain. *Mol Cell* 40 (3):455-464
- Ribeiro SA, Gatlin JC, Dong Y, Joglekar A, Cameron L, Hudson DF, Farr CJ, McEwen BF, Salmon ED, Earnshaw WC, Vagnarelli P (2009) Condensin regulates the stiffness of vertebrate centromeres. *Mol Biol Cell* 20 (9):2371-2380
- Rocchi M, Archidiacono N, Schempp W, Capozzi O, Stanyon R (2012) Centromere repositioning in mammals. *Heredity (Edinb)* 108 (1):59-67
- Rudd MK, Mays RW, Schwartz S, Willard HF (2003a) Human artificial chromosomes with alpha satellite-based de novo centromeres show increased frequency of nondisjunction and anaphase lag. *Mol Cell Biol* 23 (21):7689-7697
- Rudd MK, Schueler MG, Willard HF (2003b) Sequence organization and functional annotation of human centromeres. *Cold Spring Harb Symp Quant Biol* 68:141-149
- Rudd MK, Willard HF (2004) Analysis of the centromeric regions of the human genome assembly. *Trends Genet* 20 (11):529-533
- Santaguida S, Musacchio A (2009) The life and miracles of kinetochores. *EMBO J*
- Schueler MG, Swanson W, Thomas PJ, Green ED (2010) Adaptive evolution of foundation kinetochore proteins in primates. *Mol Biol Evol* 27 (7):1585-1597
- Scott KC, Merrett SL, Willard HF (2006) A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains. *Curr Biol* 16 (2):119-129
- Seuanez H, Fletcher J, Evans HJ, Martin DE (1976) A chromosome rearrangement in orangutan studied with Q-, C-, and G-banding techniques. *Cytogenet Cell Genet* 17 (1):26-34
- Shang WH, Hori T, Toyoda A, Kato J, Pependorf K, Sakakibara Y, Fujiyama A, Fukagawa T (2010) Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. *Genome Res* 20 (9):1219-1228

- She X, Horvath JE, Jiang Z, Liu G, Furey TS, Christ L, Clark R, Graves T, Gulden CL, Alkan C, Bailey JA, Sahinalp C, Rocchi M, Haussler D, Wilson RK, Miller W, Schwartz S, Eichler EE (2004) The structure and evolution of centromeric transition regions within the human genome. *Nature* 430 (7002):857-864
- Shi J, Wolf SE, Burke JM, Presting GG, Ross-Ibarra J, Dawe RK (2010) Widespread gene conversion in centromere cores. *PLoS Biol* 8 (3):e1000327
- Smit AFA, Riggs AD (1996) Tiggers and DNA transposon fossils in the human genome. *Proc Natl Acad Sci U S A* 93 (4):1443-1448
- Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191 (4227):528-535
- Smith KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M (2011) Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. *Mol Cell Biol* 31 (12):2528-2542
- Southern EM (1975) Long range periodicities in mouse satellite DNA. *J Mol Biol* 94 (1):51-69
- Steiner NC, Clarke L (1994) A novel epigenetic effect can alter centromere function in fission yeast. *Cell* 79 (5):865-874
- Stoyan T, Carbon J (2004) Inner kinetochore of the pathogenic yeast *Candida glabrata*. *Eukaryot Cell* 3 (5):1154-1163
- Sullivan BA, Schwartz S (1995) Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. *Hum Mol Genet* 4 (12):2189-2197
- Sun X, Wahlstrom J, Karpen G (1997) Molecular structure of a functional *Drosophila* centromere. *Cell* 91 (7):1007-1019
- Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. *J Biol* 3 (4):18

- Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *Plant Cell* 14 (5):1053-1066
- Thompson SL, Bakhoun SF, Compton DA (2010) Mechanisms of chromosomal instability. *Curr Biol* 20 (6):R285-295
- Tomonaga T, Matsushita K, Yamaguchi S, Oohashi T, Shimada H, Ochiai T, Yoda K, Nomura F (2003) Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res* 63 (13):3511-3516
- Vafa O, Shelby RD, Sullivan KF (1999) CENP-A associated complex satellite DNA in the kinetochore of the Indian muntjac. *Chromosoma* 108 (6):367-374
- Van Hooser AA, Ouspenski II, Gregson HC, Starr DA, Yen TJ, Goldberg ML, Yokomori K, Earnshaw WC, Sullivan KF, Brinkley BR (2001) Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *J Cell Sci* 114 (Pt 19):3529-3542.
- Ventura M, Mudge JM, Palumbo V, Burn S, Blennow E, Pierluigi M, Giorda R, Zuffardi O, Archidiacono N, Jackson MS, Rocchi M (2003) Neocentromeres in 15q24-26 map to duplicons which flanked an ancestral centromere in 15q25. *Genome Res* 13 (9):2059-2068
- Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blocker 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, Guerin G, Hasegawa T, Hill EW, Jurka J, Kiiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Roed KH, Ryder OA, Searle S, Skow L, Swinburne JE, Syvanen AC, Tozaki T, Valberg SJ, Vaudin M, White JR, Zody MC, Lander ES, Lindblad-Toh K (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326 (5954):865-867
- Wang H, Xu Z, Gao L, Hao B (2009) A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol Biol* 9:195

- Warburton PE (2004) Chromosomal dynamics of human neocentromere formation. *Chromosome Res* 12 (6):617-626
- Weaver BA, Cleveland DW (2007) Aneuploidy: instigator and inhibitor of tumorigenesis. *Cancer Res* 67 (21):10103-10105
- Williams BC, Murphy TD, Goldberg ML, Karpen GH (1998) Neocentromere activity of structurally acentric mini-chromosomes in *Drosophila* [In Process Citation]. *Nat Genet* 18 (1):30-37
- Yan H, Talbert PB, Lee HR, Jett J, Henikoff S, Chen F, Jiang J (2008) Intergenic locations of rice centromeric chromatin. *PLoS Biol* 6 (11):e286
- Yunis JJ, Prakash O (1982) The origin of man: a chromosomal pictorial legacy. *Science* 215 (4539):1525-1530
- Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. *Plant Cell* 14 (11):2825-2836
- Zinkowski RP, Meyne J, Brinkley BR (1991) The centromere-kinetochore complex: a repeat subunit model. *J Cell Biol* 113 (5):1091-1110





Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Kalitsis, P; Choo, KHA

Title:

The evolutionary life cycle of the resilient centromere

Date:

2012-08-01

Citation:

Kalitsis, P. & Choo, K. H. A. (2012). The evolutionary life cycle of the resilient centromere. CHROMOSOMA, 121 (4), pp.327-340. <https://doi.org/10.1007/s00412-012-0369-6>.

Persistent Link:

<http://hdl.handle.net/11343/219761>

File Description:

Accepted version