Funct Integr Genomics (2009) 9:287–292
DOI 10.1007/s10142-009-0124-0

REVIEW

Centromeres: long intergenic spaces with adaptive features

Lisa Kanizay · R. Kelly Dawe

Received: 16 February 2009 / Revised: 20 April 2009 / Accepted: 24 April 2009 / Published online: 12 May 2009

Abstract Centromeres are composed of inner kinetochore proteins, which are largely conserved across species, and repetitive DNA, which shows comparatively little sequence conservation. Due to this fundamental paradox the formation and maintenance of centromeres remains largely a mystery. However, it has become increasingly clear that a long-standing balance between epigenetic and genetic control governs the interactions of centromeric DNA and inner kinetochore proteins. The comparison of classical neocentromeres in plants, which are entirely genetic in their mode of operation, and clinical neocentromeres, which are sequence-independent, illustrates the conflict between genetics and epigenetics in regions that control their own transmission to progeny. Tandem repeat arrays present in centromeres may have an origin in meiotic drive or other selfish patterns of evolution, as is the case for the CENP-B box and CENP-B protein in human. In grasses retrotransposons have invaded centromeres to the point of complete domination, consequently breaking genetic regulation at these centromeres. The accumulation of tandem repeats and transposons causes centromeres to expand in size, effectively pushing genes to the sides and opening the centromere to ever fewer constraints on the DNA sequence. On genetic maps centromeres appear as long intergenic spaces that evolve rapidly and apparently without regard to host fitness.

L. Kanizay · R. K. Dawe (✉)
Department of Plant Biology, Miller Plant Science Bldg,
University of Georgia,
Athens, GA 30602, USA
e-mail: kelly@plantbio.uga.edu

R. K. Dawe
Department of Genetics, University of Georgia,
Athens, GA 30602, USA

Keywords Genome conflict · Epigenetic · Grass · Human

Introduction

Proper chromosome segregation during cell division is essential for the perpetuation of life. In higher eukaryotes segregation is mediated by large protein complexes known as kinetochores, which interact on one side (inner kinetochore) with the chromosomes and on the other side (outer kinetochore) with the spindle microtubules. The kinetochore serves as the protein interface between chromosomes and the spindle apparatus that guides chromosome segregation. In recent years there has been a large effort to characterize kinetochore proteins and centromeric DNA, with the goal of understanding the specifics of their interaction. Interestingly, it has been found that kinetochore proteins are well conserved across eukaryotes, but that centromeric DNA has very limited sequence similarity even within families. How can such functionally important and conserved proteins faithfully recognize seemingly random sequences of DNA?

While the base-by-base sequence of centromeres is not conserved, there are defining features that traverse species boundaries: centromeres are composed of particular forms of repetitive DNAs and they possess the ability to recruit the histone H3 variant, CENH3. Although the type(s) and amounts of repetitive DNA vary widely across eukaryotes, they are generally consistent within species and often appear to be quite specific to centromeres. However, fine scale analyses show that centromeric sequences are rarely specific to the kinetochore-binding domains; rather, they tend to spread into pericentromeric regions as well. In the absence of identifying sequence features, centromeres have come to be identified by their interaction with CENH3. The evolution of centromeric DNA and how kinetochores

recognize it remain shrouded in mystery. What exactly is required for a functional centromere? Here we propose three elements that are essential to understanding the complex nature of centromere behavior, formation, and maintenance: repetitive DNA, CENH3, and the significance of epigenetics in sequence evolution.

Part I: centromeres: DNA and associated proteins

DNA

In humans and other great apes centromeres are exclusively composed of arrays of tandem repeats which recruit the centromeric histone H3 variant, CENH3 (referred to as CENP-A in mammals), and bind tightly to a second kinetochore protein known as CENP-B (Fig. 1b). Generally, the individual units in repeat arrays fall within a 120–200 bp size range. This is thought to be of some importance because it is about the length of DNA required to wrap around the nucleosome (that contains CENH3), providing the first level of DNA packaging in chromatin. In human, the tandem repeats are known as α satellites and are 171 bp in length. Alpha satellites vary slightly in sequence from chromosome to chromosome. This slight sequence variation is interpreted as an indication that centromeric repeat arrays are capable of rapid evolution via some combination of unequal exchange, inversion, and duplication. Subvariants of centromere repeats can be conserved to some degree between species: chimps contain α satellites that share sequence similarity with some human α satellites. The sequence conservation seen between chimps and humans is most likely due to their recent divergence time, because homology deteriorates very quickly in wide species comparisons. For example, *Drosophila* centromeres are composed of a mere 4–8 bp repeat unit. Additionally, mouse minor satellite repeats, while close in size to α satellites, share only the 17 bp Cenp-B box (discussed below) (Choo 1997; Amor et al. 2004).

Plant centromeres also contain tandem arrays that are comparable to those found in animals. They fall within the same 120–200 bp size range and bind CENH3. These arrays are also presumed to be capable of rapid evolution. For instance, the copy number of maize centromeric repeat CentC varies dramatically among chromosomes and between species (Kato et al. 2004; Lamb and Birchler 2006). Also, the major *Arabidopsis* centromere repeat (called the 180 bp repeat) shows evidence of extensive local homogenization within and among *Arabidopsis* ecotypes (Hall et al. 2005). As in human, the tandem centromere repeats in plants are conserved between recently diverged species. For example, the CentC monomers of maize and *Tripsacum* (which diverged from a common ancestor about three million years ago) are indistinguishable from one another. Even the tandem arrays from the distantly related species maize and rice (roughly 60 million years separated) share a limited homology over an 80 bp region (Lee et al. 2005).

In addition to tandem arrays, some plants appear to have evolved a novel mechanism for creating subdomains within their centromeric DNA. Most grass centromeres contain two distinct types of repetitive DNA: tandem repeat arrays and centromere specific LTR retrotransposons (CR elements) (Fig. 1c). A major difference between CR elements and tandem repeat arrays is that CR elements are conserved at a much higher level than tandem arrays. For instance, the coding portion of the elements is conserved between all CR elements. CR elements are highly abundant and have been transpositionally active in recent history. They are randomly interspersed within the tandem repeat arrays; but, new elements also frequently jump into other CR elements. The relative quantities of CR elements compared to tandem repeats vary widely among centromeres of the same indi-

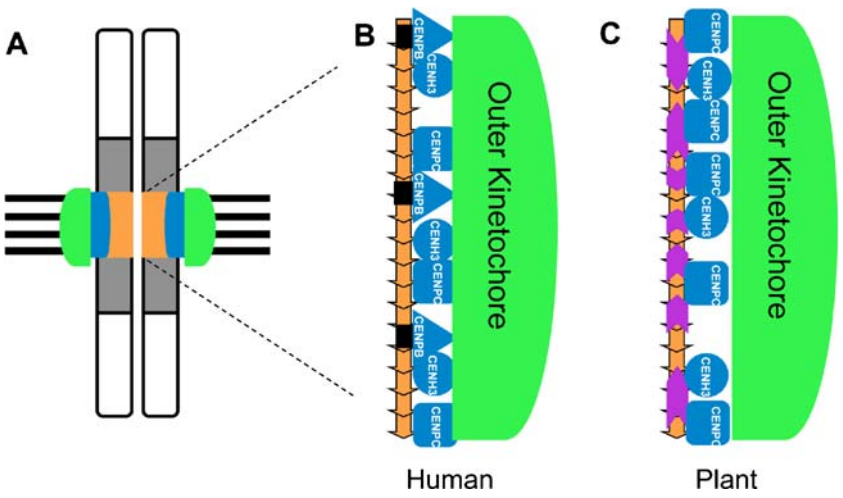

Fig. 1 Model of centromere composition in animals and plants. **a** Sister chromatids with kinetochores and spindle microtubules attached. **b** Animal centromere, with CENP-B box (*black*) in tandem arrays. **c** Plant centromere with CR elements (*purple*) and no CENP-B or CENP-B box. Centromeric DNA is *orange*, pericentromeric regions are *gray*, and inner kinetochore proteins are labeled

vidual, with CR elements being the more consistent feature. This has been clearly shown with the recent sequencing of some rice (Feng et al. 2002; Matsumoto et al. 2005) and maize (unpublished) centromeres. Due to their prolific nature CR elements are thought to be an important part of grass chromosomes. In wheat there are no known tandem repeats in centromeres and CR elements appear to have overwhelmed any other form of sequence (Liu et al. 2008).

Proteins

The next level of centromere structure is the kinetochore. This large protein complex is divided into two major domains, the inner kinetochore and the outer kinetochore. The inner kinetochore interacts with the centromeric DNA, repressing genetic recombination, while the outer kinetochore connects inner kinetochore proteins to spindle microtubules. CENH3 and CENP-C are conserved in plants and animals, and are the only two kinetochore proteins known to bind DNA in both kingdoms. Several other inner and outer kinetochore proteins have clear homologs in plants (Yu et al. 1999; Sato et al. 2005; Du and Dawe 2007); for example, Mis12, Ndc80, Nnf1, and Spc105 (Meraldi et al. 2006; Cheeseman and Desai 2008).

CENP-B is an example of an important inner kinetochore protein in animals that is absent in plants. At a glance, the lack of CENP-B may not seem particularly noteworthy, considering there are other key kinetochore proteins that have not yet been identified in plants or even in all animals. What makes CENP-B unique are its homology to the Mariner transposase and its novel function. It binds to a specific DNA sequence known as the CENP-B box, which is located in a subset of human α satellite repeats. Neither CENH3 nor CENP-C bind DNA in a sequence-dependent manner. Additionally, CENP-B is the only kinetochore protein with homology to a transposase. Interestingly, CENP-B is not essential for centromere maintenance in mammals and can be removed from mice with no adverse effects (Hudson et al. 1998; Kapoor et al. 1998; Perez-Castro et al. 1998). However, it is necessary for de novo human artificial chromosome (HAC) formation (Ohzeki et al. 2002). Functional HACs do not form efficiently over α satellites that do not contain CENP-B boxes. Plant centromeres appear to rely more heavily than their mammalian counterparts on sequence-independent mechanisms to monitor their DNA–protein interactions. We refer to these mechanisms as epigenetic. Here, and in all other studies on centromeres, the term epigenetic is used in the broadest sense. The interactions are not sequence-specific but otherwise not understood (Dawe and Henikoff 2006; Gieni et al. 2008; Birchler et al. 2009).

Clearly repetitive DNA and kinetochore proteins are integral components of both plant and animal centromeres. Because many kinetochore proteins are conserved between plants and animals, it seems paradoxical that sequence and type of DNA should vary so much between plant and animal centromeres. This discrepancy indicates that the type of repetitive DNA present in a centromere has a relatively small role in centromere function. One interpretation is that the repeats are a consequence—not a cause—of kinetochore placement (Nagaki et al. 2004; Dawe 2005).

Part II: neocentromeres: models for centromere formation

There have been cases in both plants and animals of ectopic centromere activity. Such regions that exhibit new centromere activity are termed neocentromeres and move towards the spindle poles during cell division independently of true centromeres. There are two major types, classic and clinical. These are summarized in Fig. 2. The classic type is found only in maize and rye, whereas the clinical type is found in human, barley, and was most recently discovered in an oat–maize addition line (Topp et al. in press). Each of the two categories of neocentromere exemplifies a particular aspect of true centromeres.

Classic neocentromeres

Classic neocentromeres are completely unique in composition compared to true centromeres. They employ neither the same repeats nor the kinetochore proteins utilized by true centromeres. In maize, classic neocentromeres arose as a genetic consequence of the abnormal chromosome 10 (Ab10) meiotic drive system. When they are inactive, classic neocentromeres are referred to as knobs. They are composed of very long arrays of two different tandem repeats, a 180 bp repeat and a 350 bp repeat, that form cytologically visible heterochromatic regions (Peacock et al. 1981; Dennis and Peacock 1984; Ananiev et al. 1998). When Ab10 is present in a heterozygous state, it and all other chromosomes that are heterozygous for a knob are preferentially transmitted to the female gamete (Fig. 2b). A widely accepted model for the mechanism of preferential segregation was proposed by Rhoades (1952). According to Rhoades, preferential segregation first requires a crossover between the heterozygous homologues, creating a heteromorphic dyad. The knobbed chromatids then orient to the outermost cells of what will become a linear tetrad. Following this, anaphase begins and the knobs are transformed into neocentromeres. The neocentromere repeats are believed to interact in a sequence-specific manner with the spindle, presumably through the activity of novel proteins encoded within the Ab10 haplotype (Hiatt et al. 2002; Mroczek et al. 2006). The sequence-specific interaction

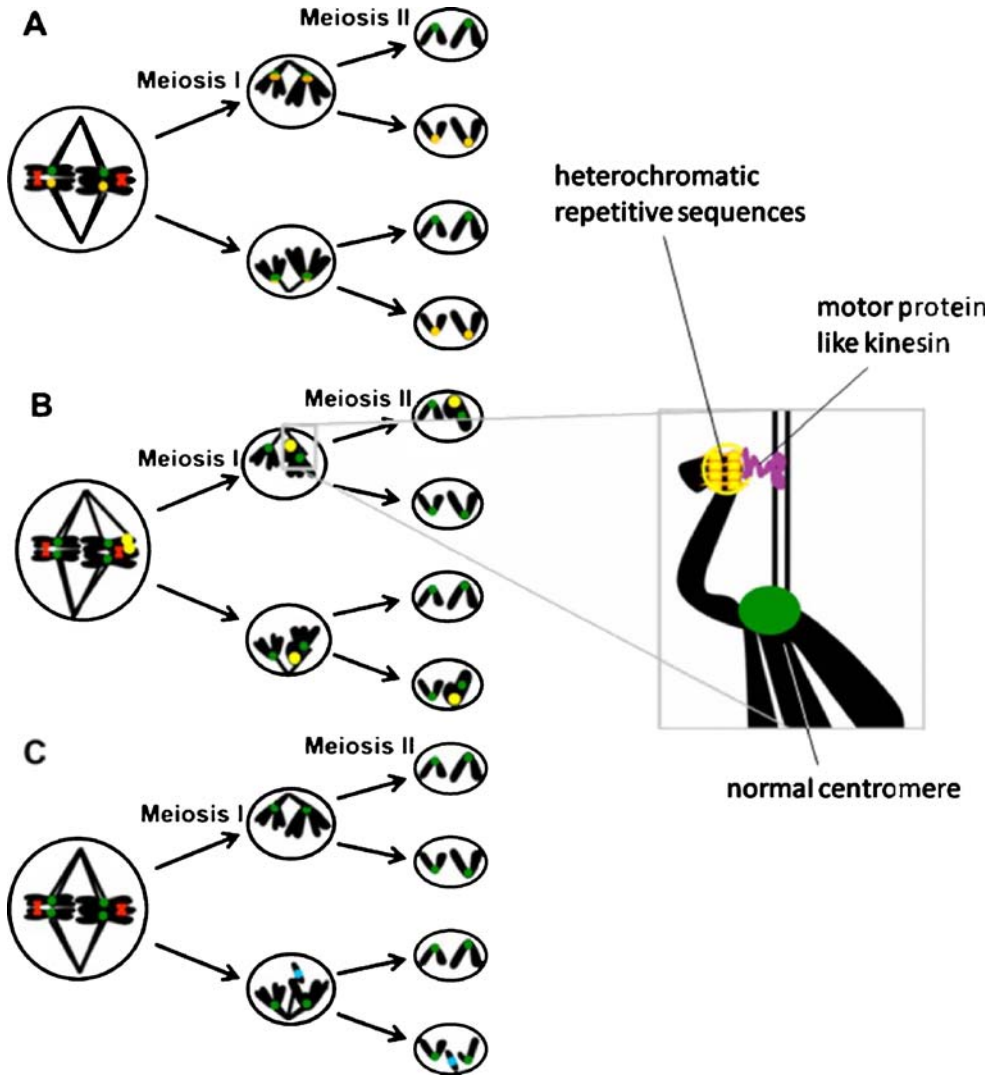

Fig. 2 Female meiosis illustrating the behavior of centromeres and neocentromeres. **a** Normal meiosis. Crossing over occurs in prophase and is complete by the time metaphase I begins. Fused sister kinetochores drive chromosomes to poles in the first reductional division. Meiosis II is much like mitosis, however, only the basal megaspore survives to produce an egg (in higher plants). The DNA that ends up in the basal megaspore will be transmitted to the next generation. **b** Classic neocentromere drive. A crossover occurs between the centromere (*green*) and knob (*yellow*), creating heteromorphic dyads. During anaphase I the knobs become neocentromeres and pull their chromosome arms to the outermost poles. This orientation is maintained through anaphase II, leaving the knobbed chromatid in the basal-most megaspore. The knobs are composed of arrays of tandem repeats that are tightly packed into heterochromatin. They move laterally along microtubules rather than end-on (like true centromeres) and are thought to interact with microtubules via a motor protein-like kinesin (*purple*, *in gray box*). **c** Clinical neocentromere formation. The process is essentially the same as normal meiosis. However, a chromosome breakage or some other aberrant event produces an arm fragment that spontaneously gains centromere function (*purple arrow*). This new centromere is often called a neocentromere (*blue*). The fragment with the neocentromere may end up in any of the four products of meiosis

allows neocentromeres to move laterally along the spindle microtubules so that they reach spindle poles before true centromeres (Yu et al. 1997). It is this fast, directed, sequence-specific movement toward the outermost spindle poles that allows neocentromere-containing chromosomes to be preferentially transmitted. A similar type of neocentromere activity has been documented in rye, although no Ab10-like locus has been discovered (Hayward 1962; Manzanero and Puertas 2003; Puertas et al. 2005).

The Ab10 system serves as an excellent example of how chromosome domains may adapt to the meiotic spindle for the selfish purpose of gaining a segregation advantage. Although it is difficult to draw clear parallels in other species, classic maize/rye neocentromeres are similar in some ways to the human alpha-satellite CENP-B interaction. Both systems seem to have evolved outside of or in parallel with the typical, epigenetically driven kinetochore–spindle interaction.

Clinical neocentromeres

Clinical neocentromeres form within the framework of the normal kinetochore and do not cause meiotic drive. They are usually observed on broken chromosomes and in euchromatic regions that contain few repeats. In spite of the complete lack of DNA structural and sequence homology between human neocentromeres and true centromeres, the two are nearly identical in kinetochore composition and both require the presence of CENH3 to function. The only discrepancy is that human neocentromeres are not able to bind CENP-B due to the lack of the CENP-B box. In plants, clinical-type neocentromeres have been found in two different chromosome addition lines, a wheat–barley addition line (Nasuda et al. 2005) and an oat–maize addition line (Topp et al., Cytogenetic and Genome Research, in press). The chromatin where the barley neocentromere formed is unknown, whereas the oat–maize neocentromere appears to have formed in euchromatin. Similar to the clinical neocentromeres in human, these plant clinical-type neocentromeres form through their ability to recruit CENH3. Clinical-type neocentromeres are a testament to the epigenetic nature of centromeres, proving that CENH3 has the potential to localize to virtually any DNA sequence. Together, classic and clinical neocentromeres illustrate the full power of the genome to adapt to the spindle in ways that allow its transmission through generations.

Part III: the mixed model for centromere function: both selfish and epigenetic inheritance

The evidence from clinical neocentromere formation indicates that the recruitment of CENH3 and other kinetochore proteins, not DNA sequence, determines the site of centromere formation. Under a sequence-independent evolution scenario it is expected that selfish DNA elements will proliferate and this is indeed the case. The retroelement group of transposable elements (TEs) is the primary pioneer of non-coding spaces, rapidly accumulating within all types of intergenic (neutral) regions. The CR elements are a particularly successful group of retroelements specific to grass centromeres. However, tandem repeats have also repeatedly arisen in centromeres—unlike any other intergenic space in the genomes of complex eukaryotes. Presumably, the success of tandem repeats in centromeres is a response to some form of genetic fitness advantage. This is thought to be an outcome of two properties: a proclivity of centromere repeats to bind inner kinetochore proteins and an ability of the repeats to access a novel mechanism for amplifying in copy number.

The one-cell-takes-all system of female gamete formation provides the perfect arena for centromere repeats to gain a genetic fitness advantage (Sandler and Novitski 1957; Henikoff et al. 2001). The egg system of both plants and animals allows for asymmetric inheritance and a capacity to proliferate independently of the DNA duplication events of S phase (Fig. 2). Particularly clear examples of this are the classic neocentromeres of maize. Here the repetitive DNA of knobs exists in two heteromorphic dyads at the start of cell division, enabling their orientation to the outermost spindle poles and thus ensuring their transmission to the next generation. The action of classic neocentromeres ensures the transmission of the repeats as well as any other selfish DNA that may be associated with them. Therefore, such a mechanism naturally benefits transposable elements that can target the repeat arrays. It stands to reason that a process similar in outcome (if not mechanism) occurs during centromere evolution (Henikoff et al. 2001). Tandem repeats may occasionally acquire sequence-specific interactions with inner kinetochore proteins and proliferate as a result (Dawe and Henikoff 2006).

Tandem repeats are particularly easy to detect and no doubt dominate in many species. However, despite their apparent abundance, the numbers and types of repeats are sporadic and inconsistent over the tree of life and within major species lineages. It appears that the fitness advantages of tandem repeats are tenuous and easily broken, either by a selected mutation (for instance as a means to restrict centromere drive) or by simple drift of kinetochore protein sequence. Once a major fitness advantage is lost, centromeres appear to rapidly take on features of a massive intergenic region and accumulate transposons. The retrotransposon laden centromeres of wheat (Liu, et al. 2008) and Neurospora (Cambareri et al. 1998) stand out in their evolutionary clades as examples where epigenetics have taken over centromeres. Transposons also appear to dominate over tandem repeats in the marsupial lineage (Gentles et al. 2007). However, any new repeat can, over time, transition to a more genetic interaction by mutational events that confer a fitness advantage to the overlying kinetochore.

Taken together, the available data suggest that in most cases kinetochores are targeted to DNA sequences epigenetically without regard to sequence. Once a domain is occupied by a kinetochore, it is shielded from crossing over and loses many of the normal constraints on DNA evolution, particularly with regard to repetitive DNA. Simple repeat arrays often evolve within centromeres and can confer minor fitness advantages to associated kinetochores such that they proliferate in a manner similar to meiotic drive. In some lineages, tandem repeats are the minor component and specialized retrotransposons have evolved to fill the centromere niche. One result is that the space occupied by centromeres tends to expand over time. On genetic maps centromeres appear as long intergenic spaces with a great abundance of unusual repeats that change rapidly.

References

Amor DJ, Kalitsis P, Sumer H, Choo KHA (2004) Building the centromere: from foundation proteins to 3D organization. Trends Cell Biol 14:359–368

Ananiev EV, Phillips RL, Rines HW (1998) A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? Proc Natl Acad Sci U S A 95:10785–10790

Birchler J, Gao Z, Han F (2009) A tale of two centromeres—diversity of structure but conservation of function in plants and animals. Funct Integr Genomics 9:7–13

Cambareri EB, Aisner R, Carbon J (1998) Structure of the chromosome VII centromere region in neurospora crassa: degenerate transposons and simple repeats. Mol Cell Biol 18:5465–5477

Cheeseman IM, Desai A (2008) Molecular architecture of the kinetochore-microtubule interface. Nat Rev Mol Cell Biol 9:33–46

Choo KHA (1997) The centromere. Oxford University Press, Oxford

Dawe RK (2005) Centromere renewal and replacement in the plant kingdom. Proc Natl Acad Sci U S A 102:11573–11574

Dawe RK, Henikoff S (2006) Centromeres put epigenetics in the driver's seat. Trends Biochem Sci 31:662–669

Dennis ES, Peacock WJ (1984) Knob heterochromatin homology in maize and its relatives. J Mol Evol 20:341–350

Du Y, Dawe RK (2007) Maize NDC80 is a constitutive feature of the central kinetochore. Chromosome Res 15:767–775

Feng Q, Zhang YJ, Hao P, Wang SY et al (2002) Sequence and analysis of rice chromosome 4. Nature 420:316–320

Gentles AJ, Wakefield MJ, Kohany O, Gu W et al (2007) Evolutionary dynamics of transposable elements in the short-tailed opossum Monodelphis domestica. Genome Res 17:992–1004

Gieni RS, Chan GKT, Hendzel MJ (2008) Epigenetics regulate centromere formation and kinetochore function. J Cell Biochem 104:2027–2039

Hall SE, Luo S, Hall AE, Preuss D (2005) Differential rates of local and global homogenization in centromere satellites from *Arabidopsis* relatives. Genetics 170:1913–1927

Hayward MD (1962) Genetic control of neocentric activity in rye. Heredity 17:439–441

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

Hiatt EN, Kentner EK, Dawe RK (2002) Independently regulated neocentromere activity of two classes of tandem repeat arrays. Plant Cell 14:407–420

Hudson DF, Fowler KJ, Earle E, Saffery R et al (1998) Centromere protein B null mice are mitotically and meiotically normal but have lower body and testis weights. J Cell Biol 141:309–319

Kapoor M, Luna RMD, Liu G, Lozano G et al (1998) The cenpB gene is not essential in mice. Chromosoma 107:570–576

Kato A, Lamb J, Birchler J (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci U S A 101:13554–13559

Lamb JC, Birchler JA (2006) Retroelement genome painting: cytological visualization of retroelement expansions in the genera Zea and Tripsacum. Genetics 173:1007–1021

Lee HR, Zhang WL, Langdon T, Jin WW et al (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in Oryza species. Proc Natl Acad Sci U S A 102:11793–11798

Liu Z, Yue W, Li DY, Wang RRC et al (2008) Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres. Chromosoma 117:445–456

Manzanero S, Puertas MJ (2003) Rye terminal neocentromeres: characterization of the underlying DNA and chromatin structure. Chromosoma 111:408–415

Matsumoto T, Wu JZ, Kanamori H, Katayose Y et al (2005) The map-based sequence of the rice genome. Nature 436:793–800

Meraldi P, McAinsh AD, Rheinbay E, Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. Genome Biology 7:R23

Mroczek RJ, Melo JR, Luce AC, Hiatt EN, Dawe RK (2006) The maize Ab 10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. Genetics 174:145–154

Nagaki K, Cheng ZK, Ouyang S, Talbert PB et al (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Nasuda S, Hudakova S, Schubert I, Houben A, Endo TR (2005) Stable barley chromosomes without centromeric repeats. Proc Natl Acad Sci U S A 102:9842–9847

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 159:765–775

Peacock WJ, Dennis ES, Rhoades MM, Pryor AJ (1981) Highly repeated DNA-sequence limited to knob heterochromatin in maize. Proc Natl Acad Sci U S A 78:4490–4494

Perez-Castro AV, Shamanski FL, Meneses JJ, Lovato TL et al (1998) Centromeric protein B null mice are viable with no apparent abnormalities. Dev Biol 201:135–143

Puertas MJ, Garcia-Chico R, Sotillo E, Gonzalez-Sanchez M, Manzanero S (2005) Movement ability of rye terminal neo-centromeres. Cytogenetic Genome Res 109:120–127

Rhoades, MM (1952) Preferential segregation in maize, pp. 66–80 in *Heterosis*, edited by J. W. Gowen. Iowa State College Press, Ames, Iowa

Sandler L, Novitski E (1957) Meiotic drive as an evolutionary force. Am Nat 91:105–110

Sato H, Shibata F, Murata M (2005) Characterization of a Mis12 homologue in *Arabidopsis thaliana*. Chromosome Res 13:827–834

Yu HG, Hiatt EN, Chan A, Sweeney M, Dawe RK (1997) Neocentromere-mediated chromosome movement in maize. J Cell Biol 139:831–840

Yu HG, Muszynski MG, Dawe RK (1999) The maize homologue of the cell cycle checkpoint protein MAD2 reveals kinetochore substructure and contrasting mitotic and meiotic localization patterns. J Cell Biol 145:425–435