Centromeres in the genomic era: unraveling paradoxes
Gregory P Copenhaver* and Daphne Preuss

The centromeres of higher plants and animals share many common features, though current models fail to account for all aspects of centromere composition and function. This dilemma is likely to be resolved in the next few years in Arabidopsis where robust assays for centromere function are available and the sequence of the entire genome will be determined.

Addresses
University of Chicago, 1103 E. 57th Street, Chicago, IL 60637, USA
*e-mail: gpcopenh@midway.uchicago.edu

http://biomednet.com/elecref/1369526600200104
© Elsevier Science Ltd ISSN 1369-5266

Introduction
Most eukaryotic chromosomes contain a specialized domain — the centromere — that mediates critical mitotic and meiotic functions, including kinetochore nucleation, spindle attachment, and sister chromatid cohesion. These processes ensure faithful chromosome partitioning during cell division. Intense scrutiny of the DNA and proteins that comprise the centromere revealed its fundamental roles in chromosome mechanics and, in some cases, enabled the creation of artificial chromosomes. The centromeric DNA of several lower eukaryotes has been completely defined; much less is known about the centromeric DNA of higher eukaryotes, including plants. In this review we define parallels with lower eukaryotic systems, survey current models for higher eukaryotic centromere function, suggest guidelines for assessing models, and summarize the short-term prospects for investigating plant centromeres.

Centromeres from lower eukaryotes
The centromeres from many fungal organisms were mapped by tetrad analysis, a powerful genetic tool used to map sites of genetic exchange relative to the centromere (Figure 1) [1]. Tetrad analysis is possible when meiotically-related daughter cells can be recovered and analyzed; yeasts such as Saccharomyces cerevisiae and Schizosaccharomyces pombe package their meiotic products into asci that can be readily separated and characterized. Using tetrad analysis and chromosome walking, Clarke and Carbon [2] identified the centromere of S. cerevisiae chromosome III, mapping the chromosomal regions that attach to the spindle and migrate to opposite poles in meiosis I, thus providing a functional definition of the centromere. Chromosome walking delineated a small region of ~20 kb, and subsequently, artificial chromosomes were used to show that only 125 bp were needed for centromere function [3]. A highly conserved copy of this sequence resides at the centromere of every S. cerevisiae chromosome. This minimal region meets two stringent criteria: first, artificial chromosomes containing the 125 bp region segregate efficiently through mitosis, with loss rates approaching those of normal chromosomes (>99% transmission in the absence of selection); and second, these constructs partition efficiently through meiosis, as assayed by tetrad analysis [4]. Similar genetic analyses led to the elucidation of the S. pombe centromere [5]. Chromosome walking on all three chromosomes pinpointed regions ranging from 40 to 100 kb. Unlike S. cerevisiae centromeres, each S. pombe centromere contains a divergent core region surrounded by repetitive elements. S. pombe artificial chromosomes define a minimal centromere region that confers high-fidelity transmission: this ~7 kb fragment consists of a portion of the chromosome II centromere core and a single flanking repeat element.

Centromeres from higher eukaryotes
Tetrad analysis has not been used to map the centromeres of most higher eukaryotes because the products of individual meioses are relatively difficult to obtain and analyze. This limitation has hampered progress toward genetically defining these centromeres in their native contexts. In an alternative strategy, the DNA sequences that comprise higher eukaryotic centromeres are typically localized genetically by analyzing the assortment of chromosome fragments and rearrangements. These techniques have limited resolution — breakpoints close to the centromeres cannot always be obtained, and the activity of centromeres can be altered when they are removed from their natural context. Nonetheless, such methods have localized centromere functions to regions ranging from a few hundred kilobases (Drosophila and maize) to several megabases (tomato, rice, and humans) [6••,7–11]. The relatively large chromosome size typical of these organisms facilitates cytological analysis of centromeres, and the genetically defined regions typically correlate with the highly condensed chromatin known as the centromeric constriction [12,13•]. These studies have been the subject of several recent reviews [14–17]. The emerging trend is that the large regions that encompass higher eukaryotic centromeres resemble the centromeres defined in S. pombe. All higher organisms examined thus far, including plants, contain distinctive and numerous repetitive elements at each centromere; other than these repeats, a conserved sequence that resides at the centromere of each chromosome has not been found [15,18,19•].

Genomic sequencing is underway in several multicellular organisms, promising to yield a dramatically improved understanding of centromeric DNA in the coming years. The first complete sequence of a higher eukaryote, that of the nematode C. elegans, unfortunately does not clarify the understanding of centromeres: C. elegans has holocentric chromosomes that attach to spindle fibers along their
entire length [20]. In contrast, the complete sequence of Arabidopsis thaliana, an organism with typical centromeres, will be obtained within the next two years [21]. Arabidopsis is particularly suited for elucidating centromere functions — among all multicellular model organisms, it is uniquely amenable to tetrad analysis. The Arabidopsis quartet mutation causes the four products of male meiosis to remain attached, yielding tetrads of pollen grains [22]. Recently, tetrad analysis was used to genetically map all five centromeres in Arabidopsis [23**]; high resolution genetic mapping will allow a precise definition of the DNA sequences that confer centromere function. Tetrad analysis in combination with sequence information will clarify whether centromeres consist entirely of repeats or if a unique centromeric element resides within the repetitive region.

**Centromere paradoxes**

Neither the DNA sequences that comprise higher eukaryotic centromeres, nor the structure of their centromeric chromatin are fully understood; consequently, only conjectural models for centromere function are available. Appropriate explanations must account for several puzzling observations, including the diversity of centromere DNA sequences, the evolutionary conservation of some centromere binding proteins, and the accumulation of repetitive DNA in the centromere regions.

In many higher eukaryotes, arrays of repetitive DNA measuring approximately 170 bp have been identified at the centromere, yet the sequence of these repeats is typically species specific. The centromeres from the budding yeasts Kluyveromyces lactis, Kluyveromyces marxianus, Candida glabrata, and S. cerevisiae are more similar in size and sequence composition, yet none function efficiently when introduced into S. cerevisiae or S. pombe artificial chromosomes [1]. By contrast, the proteins that bind to centromeres display remarkable conservation between species. For example, homologs of the human centromere binding proteins CENP-A, CENP-B, and CENP-C have been identified in S. cerevisiae or S. pombe. CENP-A encodes a modified histone H3 protein, and CENP-B, which binds to repetitive arrays, shares sequence similarity with transposase [1]. The diverse nature of centromere DNA sequences in higher eukaryotes together with the observed homology among centromere binding proteins support the idea that the conservation of centromere functions derives from a secondary or tertiary structure, rather than a particular DNA sequence. The lack of a complete DNA sequence from any higher eukaryotic centromere, however, leaves open the possibility that a conserved sequence nucleates centromere formation in multicellular organisms.

The ubiquity of repetitive elements at the centromeres of higher eukaryotes led to the suggestion that repeats are required for centromere function. These repetitive arrays have been postulated to form essential higher order structures, bind key centromeric proteins, or serve as targets for critical DNA modification. Support for the requirement for repetitive centromeric DNA has come from the construction of human artificial chromosomes [24**,25**]. Constructs containing long arrays of the human centromere associated repeat known as α satellite DNA together with telomeres acquire modest mitotic stability in cell lines. However, these constructs often acquire other sequences in vivo, raising the possibility that α satellite DNA alone is insufficient to confer centromere function. Importantly, the role of repetitive DNA has been called into question by the recent characterization of highly stable human neocentromeres — centromeric constrictions that assemble at an atypical site on the chromosome. Many of these neocentromeres contain neither α satellite DNA nor display associated CENP-B protein [26**]. Recent findings raise the possibility that centromeric repeats are not required for centromere functions, but rather that the intrinsic nature of the centromeric regions, particularly the relatively low level of meiotic recombination, may lead to the accumulation of repetitive DNA.

**Models for centromere function in higher eukaryotes**

In multicellular organisms, the similarities between centromere functions and the paradoxes still remaining have lead to the generation of several competing models that aim to explain centromere activity (Figure 2). These models can be grouped into two categories: firstly, sequence-based models that claim higher centromeres, like those from S. cerevisiae, depend on a specific sequence, but that this sequence may be dispersed or appear in

---

**Figure 1**

A tetrad of meiotically related cells is formed following meiosis I (MI) and meiosis II (MII). Genetic analysis of marker assortment in tetrads can define sites where paternal (white chromosomes) and maternal (hatched chromosomes) DNA always migrate to opposite poles. These regions constitute a genetically defined centromere.
Models for centromere function in higher eukaryotes: (a) a unique sequence present at a single location on each chromosome; (b) dispersed centromere interrupted by nonfunctional DNA; (c) fully functional elements represented multiple times across the centromere region; (d) higher order DNA structures, DNA modification, or association with specific chromatin elements establishes a specialized domain subject to epigenetic regulation; and (e) DNA in the centromere region passively excludes an activity, thereby establishing a specialized domain recognized by the cell as a centromere.

multiple copies along the centromeric region, and secondly, sequence-independent models that invoke epigenetic modification of centromeric DNA, higher order DNA structures, or an entirely passive role for sequences that comprise the centromere.

Although the paradigms established in lower eukaryotes are appealing, they have yet to be demonstrated in more complex systems. Currently, the most complete understanding of higher centromeres comes from extensive characterization of Drosophila mini-chromosomes [6**]. These studies argue that Drosophila centromeres may be composed entirely of repetitive DNA. Final proof of this possibility will be achieved after complete DNA analysis rules out the presence of small islands of unique sequence. In addition, because these minichromosomes have undergone significant rearrangements, they may reflect neocentromeric rather than native centromeric activity. Nonetheless, this body of work demonstrates strong distinctions between higher and lower eukaryotic centromeres context.

If higher eukaryotes do rely on particular DNA sequences for centromere function, then those sequences may span a relatively large portion of the chromosome. The constricted regions of higher chromosomes clearly encompass megabases of DNA, with multiple spindle fiber attachments during cell division. In some cases, for example the maize B chromosomes, it is possible to split a centromere, yielding two chromosome fragments that segregate normally [10]. These observations suggest that critical DNA sequences may be spread throughout the centromeric region, with the space between such elements filled by apparently nonfunctional DNA.

Because a unique centromere sequence has not been demonstrated in higher eukaryotes, several recent models focus on non-sequence based mechanisms. Such models are appealing and could explain why centromeres moved from their native host into another eukaryotic species are stable, yet centromeres transferred first through E. coli can rarely function in a heterologous species. Two strong examples of epigenetic regulation have been demonstrated in Drosophila and S. pombe. Fragments of Drosophila chromosomes that are normally telomeric can acquire modest segregation stability after residing near a centromere [27•]. Control fragments that have never resided near a centromere do not share this property. These observations led to the proposal that centromere activity can spread to adjacent DNA through an unknown epigenetic mechanism. Although these results are intriguing, the relatively low transmission rates (14–24%) indicate that complete stability is not acquired by these fragments. Epigenetic control of centromere function in S. pombe has been proposed following the quantitative analysis of artificial chromosome transmission [28]. Curiously, a construct bearing a partial centromere was transmitted with a high fidelity in some colonies and a low fidelity in others. This difference in transmission frequency was not due to mutation, and rapid switching between states, through an unknown mechanism, was observed. These findings are quite unusual for yeast artificial chromosomes and reveal a potentially exciting level of epigenetic control; however, only partial centromeres display this activity, and inclusion of the complete centromeric sequence always resulted in highly stable transmission ($5 \times 10^{-4}$). Thus, for both Drosophila and S. pombe, the proposed epigenetic regulation may be most critical in cases where only a partial centromere is present.

Recently, Csink and Henikoff [29] proposed an alternative model in which centromeric DNA excludes a specific activity, such as the DNA replication machinery, and thus creates an unusual region that de facto becomes a centromere. In this model, the sequence of the region is not important as long as it does not participate in the critical activity—in this example, replication early in the cell cycle. This model requires no sequence specificity, could account for the abundance of repetitive centromeric DNA, and can be experimentally tested by manipulating reagents that are currently available. There is a vast amount of experimental evidence, however, that indicates large chromosome fragments that lack a centromere from the wheat, maize and human genomes are not segregated efficiently, despite the expectation that they would contain sizable regions of late-replicating or repetitive DNA. Consequently, this model must be modified to account for
the observation that only certain repetitive regions possess the necessary qualities to create a centromere, whereas other neocentromeric regions that lack repetitive arrays can function as efficient centromeres.

Conclusions

Many aspects of higher eukaryotic centromere function promise to be elucidated in the coming years. It will be interesting to explain features uncommon in lower eukaryotes, including the presence of large arrays of repetitive DNA, the dispersal of centromere functions along a large chromosomal region, and the potential for epigenetic regulation. Attempts to meet these challenges require rigorous criteria, like the elegant assays performed in lower eukaryotes, for assessing centromere function. In particular, a thorough understanding of centromeres requires quantitative genetic tests that measure chromosome transmission per cell division, a complete understanding of DNA sequences surrounding the centromere in its native context, and ideally, the development of artificial chromosomes that can reconstitute centromere activity.

A quantitative distinction between complete centromere activity and partial centromere function must be established. A fully functional centromere is expected to segregate, in the absence of selection, with fidelity approaching a native chromosome. Moreover, it should also display stability through both the first and second divisions of meiosis, reflecting proper sister chromatid cohesion and release. Cell autonomous color markers make it possible to achieve these goals, even in a multicellular organism. Like the colony color assays developed by Heiter and colleagues [30], cell autonomous assays that rely on green fluorescent protein or other vital dyes will enable chromosome transmission to be studied in plants and animals. With such assays, chromosome loss rates per cell division can be determined; these calculations will depend on monitoring cell lineage either in intact tissues with defined patterns of division or in cells cultured on a solid substrate.

The vast genomic sequencing effort will have a profound impact on the understanding of centromere functions in plants and animals. First, the availability of a complete genome sequence, along with robust genetic assays for centromere function, will allow an assessment of the role of repetitive, low copy or unique centromeric sequences. As functional genomic assays are developed, such as DNA microarrays, it will become feasible to identify deletions and insertions that may alter centromere function, as well as to characterize the expression of all of the genes in the centromeric region. The rapid advances in genomic technology, the availability of a powerful genetic assay for centromere function in Arabidopsis, and the potential for developing artificial chromosomes in higher eukaryotes promise to yield a much improved understanding of higher centromeres in the next decade.

Acknowledgements

We thank Jake Mayfield for critical reading of the manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- • of outstanding interest


This paper provides the first detailed description of a complex eukaryotic centromere. From deletion derivatives of the X chromosome, a 420 kb region was inferred to be capable of supporting centromere function, although with less efficiency than native centromeres. Importantly, after careful DNA hybridization analyses, this region was found to be composed primarily of repetitive DNA and retroelements. Because these sequences are neither centromere specific nor found an all chromosomes, a non-sequence specific basis for centromere function was suggested.


This study highlights the power of cytogenetics in Arabidopsis, showing that FISH analysis can be used to map and determine the size of DNA elements, including those that reside in centromeric heterochromatin.


An important paper which used pulsed field gel technology to genetically map large arrays of the 180 bp centromere associated sequence pAL1 in *Arabidopsis*. This paper demonstrates that the 180 bp repeats positions correlate with the position of the functional centromeres defined in [23••].


This work demonstrates the first use of tetrad analysis in *Arabidopsis* using the quartet mutation. Tetrad analysis was used to functionally define the regions containing all five *Arabidopsis* centromeres and determine their location on the genetic map.


This paper provides compelling evidence for an epigenetic model of centromere function. DNA not normally associated with the centromere was shown to acquire centromeric activity in higher eukaryotes after being near a native centromere.


This ground breaking work illustrated the feasibility of reconstituting centromere function from cloned fragments of DNA including large arrays of α satellite DNA.


This careful analysis reveals that the neocentromeric DNA of a naturally occurring human minichromosome does not contain detectable amounts of α satellite sequences despite the fact that it binds the centromere associated proteins CENP-A and CENP-C. These findings challenge the idea that repetitive DNA is necessary for centromere function.


This paper provides compelling evidence for an epigenetic model of centromere function. DNA not normally associated with the centromere was shown to acquire centromeric activity in higher eukaryotes after being near a native centromere.

