The rapidly evolving field of plant centromeres
Anne E Hall\textsuperscript{1}, Kevin C Keith\textsuperscript{1}, Sarah E Hall\textsuperscript{2}, Gregory P Copenhaver\textsuperscript{3} and Daphne Preuss\textsuperscript{1,4}

Meiotic and mitotic chromosome segregation are highly conserved in eukaryotic organisms, yet centromeres — the chromosomal sites that mediate segregation — evolve extremely rapidly. Plant centromeres have DNA elements that are shared across species, yet they diverge rapidly through large- and small-scale changes. Over evolutionary time-scales, centromeres migrate to non-centromeric regions and, in plants, heterochromatin knobs can acquire centromere activity. Discerning the functional significance of these changes will require comparative analyses of closely related species. Combined with functional assays, continued efforts in plant genomics will uncover key DNA elements that allow centromeres to retain their role in chromosome segregation while allowing rapid evolution.

Addresses
\textsuperscript{1}Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, Illinois 60637, USA
\textsuperscript{2}Committee on Genetics, The University of Chicago, Chicago, Illinois 60637, USA
\textsuperscript{3}Department of Biology and the Carolina Center for Genome Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA
\textsuperscript{4}Howard Hughes Medical Institute, The University of Chicago, Chicago, Illinois 60637, USA
\textsuperscript{4}e-mail: dpreuss@midway.uchicago.edu

This review comes from a themed issue on Genome studies and molecular genetics
Edited by Joseph R Ecker and Doug Cook
DOI 10.1016/j.pbi.2004.01.008

Plant centromere DNA content
Satellite and middle repetitive centromere DNA has been characterized in several plant species (Table 1). These sequences provide probes for identifying centromere clones from large-insert genomic libraries. Such efforts, along with whole-genome shotgun sequencing, are generating an extensive database of centromere sequences, particularly from the grasses and crucifers. The centromeres of Arabidopsis are the most thoroughly characterized, with assembled sequences extending from the chromosome arms to the satellites of all five chromosomes [1,2]. Arabidopsis centromeres contain 2.8–4 Mb tracts of tandemly repeated 178 bp satellites, occasionally interrupted by insertions of Athila, a Ty3/gypsy retroelement [1–5]. The middle repetitive regions that flank the satellites also contain Athila and other retro-elements, along with 5S ribosomal DNA (rDNA) tracts, transposable elements, and pseudogenes [1–5]. Despite the heterochromatic environment, the middle repetitive regions contain more than 200 expressed genes, at a density that is about one tenth of that on the chromosome arms [2]. Intriguingly, tracts of hemi-methylated cytosine residues are present throughout the Arabidopsis centromeres [6]. Whether these modifications contribute to centromere function or result from the impact of centromere structure on DNA-methylation mechanisms remains unresolved.
Although efforts to sequence the rice and maize centromeres are not complete, they have yielded intriguing data that are useful for comparative analysis. In rice, a contig of 1.16 Mb from centromere 4 (CEN4) revealed tracts of the CentO satellite (155 bp or 164 bp) interspersed with the Ty3/gypsy retroelement CRR (for centromere-specific retrotransposon of rice), as well as flanking regions that are rich in transposons, retroelements, and pseudogenes [7,8]. In maize, two sequenced centromere bacterial artificial chromosomes (BACs) consist largely of CentC satellite (156 bp) and the centromere-specific retrotransposon of maize (CRM), a retroelement with homology to CRR [9,10]. CentO and CentC are similar, yet neither shows homology to Arabidopsis satellites [8]. In maize, heterochromatic knobs that can serve as alternatives to the conventional centromeres have also been characterized. These knobs contain two satellites, measuring 180 bp and 350 bp, that differ from CentC [11]. Supernumerary maize

Table 1

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sequence type a</th>
<th>Size of major satellite</th>
<th>Example of recent work b</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>AS</td>
<td>171 bp</td>
<td>365 kb of Chromosome X sequenced</td>
<td>[45]</td>
</tr>
<tr>
<td>Mouse</td>
<td>AS</td>
<td>Major 234 bp, minor 120 bp</td>
<td>Examined evolution of pericentric DNA</td>
<td>[41-5]</td>
</tr>
<tr>
<td>Fruit fly</td>
<td>AS</td>
<td>5 bp</td>
<td>80 kb of CEN sequenced</td>
<td>[46]</td>
</tr>
<tr>
<td>Brassicaceae family</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabidopsis thaliana</td>
<td>AS</td>
<td>178 bp</td>
<td>Nearly complete sequence for all five CENS</td>
<td>[1–5]</td>
</tr>
<tr>
<td>Arabidopsis arenosa</td>
<td>US</td>
<td>178 bp</td>
<td>Satellites mapped by FISH</td>
<td>[47]</td>
</tr>
<tr>
<td>Arabidopsis suecica</td>
<td>US</td>
<td>178 bp</td>
<td>Satellites mapped by FISH</td>
<td>[47]</td>
</tr>
<tr>
<td>Olimarabidopsis pumila</td>
<td>US</td>
<td>168 bp, 178 bp</td>
<td>Satellites mapped by FISH</td>
<td>[19]</td>
</tr>
<tr>
<td>Brassica napus</td>
<td>US</td>
<td>176 bp</td>
<td>Satellites mapped by FISH</td>
<td>[48]</td>
</tr>
<tr>
<td>Crops</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maize</td>
<td>AS</td>
<td>156 bp</td>
<td>Two CEN BACs sequenced</td>
<td>[9*]</td>
</tr>
<tr>
<td>Rice</td>
<td>AS</td>
<td>155 bp</td>
<td>1.16 Mb contig of CEN4 sequenced</td>
<td>[7]</td>
</tr>
<tr>
<td>Barley</td>
<td>US</td>
<td>Not known</td>
<td>18 kb of BAC clone sequenced</td>
<td>[49]</td>
</tr>
<tr>
<td>Beet</td>
<td>US</td>
<td>158–160 bp</td>
<td>Satellites/transposons mapped by FISH</td>
<td>[50]</td>
</tr>
<tr>
<td>Soybean</td>
<td>US</td>
<td>120 bp</td>
<td>Satellites mapped by FISH</td>
<td>[51]</td>
</tr>
<tr>
<td>Sorghum</td>
<td>US</td>
<td>140 bp</td>
<td>Satellites mapped by FISH</td>
<td>[52]</td>
</tr>
<tr>
<td>Radish</td>
<td>US</td>
<td>177 bp</td>
<td>Satellites mapped by FISH</td>
<td>[53]</td>
</tr>
<tr>
<td>Wheat</td>
<td>US</td>
<td>537 bp</td>
<td>Satellites/transposons mapped by FISH</td>
<td>[54]</td>
</tr>
</tbody>
</table>

aAS, assembled sequence contigs; US, unassembled sequence. These sequences were predominantly generated by restriction digestions or PCR with degenerate primers. bExamples of recent work indicate the methodology employed and the type of data generated for the listed organism.
B chromosomes have a third type of centromere, containing a 1.4 kb tandemly repeated sequence with homology to knobs and a region from CEN4 [12,13]. Discerning the role of the different satellite classes and the effects of interpersed DNA elements will require functional assays.

**Centromere evolution**

Cytological and sequence analyses of centromeres from different *Arabidopsis* ecotypes have illustrated their dynamic nature; significant large-scale changes that do not disrupt centromere function have occurred over timescales of just a few million years. For example, several tandem copies of the mitochondrial genome have been integrated into CEN2 [14], an inversion involving CEN4 produced a heterochromatic knob and moved a gene-rich region into the heterochromatin [15,16], and a tract of 5S rDNA was inserted into CEN3 [17]. Whether such changes generally characterize the evolution of plant centromeres remains to be determined, but their occurrence on three of the five *Arabidopsis* centromeres warrants further analysis. Furthermore, although *Arabidopsis* chromosome arms have large tracts of DNA that reflect ancient genome duplications, similar segments have not been found in most of the centromeres, suggesting that evolutionary mechanisms differ between euchromatic regions and the centromere [1].

In addition to undergoing large-scale rearrangements, centromeres also evolve rapidly at the nucleotide level. Satellites undergo rapid change even within a species, and ecotype-specific satellite variants have been identified in *Arabidopsis thaliana* [18]. Interestingly, the nucleotide substitutions that define variants are non-random, producing highly conserved and variable regions. Although such regions imply that the nucleotide substitutions have functional significance, investigations of satellites from the closely related species *Arabidopsis arenosa* and *Olmivarabidopsis pumila* did not reveal conserved regions within these closely related genera [19]. By contrast, CentO and CentC maintain similar $5'$ and $3'$ ends despite a divergence of 50–70 million years [8,20]. In maize, analysis of satellite polymorphisms made it possible to define at least 18 different satellite classes that are arranged in higher-order repeat arrays [9]. Individual BACs contain a subset of the different satellite variants, indicating that satellites undergo local variation that is distinct from variation within the genome as a whole [9,21]. This finding is consistent with observations of human centromeres, in which satellite variants tend to cluster non-randomly because of an intrachromosomal gene-conversion mechanism [21].

Several hypotheses have surfaced to explain how a genome can maintain satellite homogeneity while still allowing rapid evolution [22]. The library hypothesis predicts that genomes contain sets of satellite variants in differing abundance. New satellites continually arise by mutation and are propagated within a genome by genetic drift. These satellites can be homogenized through unequal crossover, which would provide an explanation for chromosome-specific variation and the formation of higher-order satellite arrays [23–27]. Another model suggests that satellite evolution is driven by the selection and coevolution of satellites and centromere-binding proteins, rather than by random genetic drift. This requires that certain centromere satellite variants confer a selective advantage upon meiotic cells (i.e. eggs and sperm), a process known as meiotic drive. Such positive selection could result from a preferential interaction with centromere protein-A (CENP-A), a histone H3 variant that is incorporated into the nucleosomes that specifically bind to centromere DNA in plants, humans, yeast, and *Drosophila* [28]. If this is the case, even a slight advantage in satellite–CENP-A interactions could quickly result in genomic fixation of satellite arrays. The rapid adaptive evolution of CENP-A observed in both *Drosophila* and *Arabidopsis* is consistent with this interpretation [29,30].

Meiotic recombination is dramatically suppressed in centromere regions; the effects of this suppression on centromere DNA content and evolution are not yet clear. Mathematical models predict that low levels of homologous recombination cause the accumulation of repetitive elements [31,32], but a study that considered only non-coding *Arabidopsis* DNA found no correlation between meiotic recombination frequencies and the distribution of most transposable elements [33]. Importantly, two superfamilies of gypsy-like elements (*Athila* and *CACTA*) were excluded from the study because of their clustering within and around centromeres. On the basis of this work, Wright et al. [33] suggested that the apparent centromeric abundance of transposons is not due to suppressed recombination but rather results from ample non-coding DNA targets; insertions into chromosome arms, by contrast, are often deleterious because of the abundance of genes. Although there are few homologous recombination events in centromere regions, ectopic recombination may play a role in the evolution of these regions. For example, evidence from human and *Arabidopsis* centromere regions shows that ectopic recombination occurs at a relatively high frequency [34]. Interestingly, the centromere regions of wheat exhibit fewer local duplications than occur in the chromosome arms, resulting in a preservation of micro-synteny [35].

**Acquisition of centromere function**

Chromosome-wide studies of gene order have shown that non-centromeric regions can acquire centromere activity de novo (Figure 2a). Fluorescent in situ hybridization (FISH) mapping of closely related genomes shows that, in some cases, gene marker order remains constant while centromere positions migrate. This is the case when the X chromosomes of humans and lemur species are compared [36], and when the phylogeny of primate chromosome 6 is
reconstructed [37**]. In both cases, marker order was conserved among multiple species, and neither transposition nor inversion events could account for centromere relocation. The migration of centromeres over evolutionary time scales could result from neocentromere formation — a process whereby ectopic centromeres are activated in previously non-centromeric locations (Figure 2a). Neocentromere formation in human cell lines is triggered by the loss of an endogenous centromere, coupled with a strong selection for inheritance of the DNA fragment (Figure 2a). These new centromeres often lack satellite DNA, suggesting that associated proteins and other epigenetic factors may be more important for centromere activity than the primary DNA sequence [38].

Although the topic of centromere emergence has not been addressed in most studies of plant genome synteny, satellite-rich heterochromatic knobs can often function as...
neocentromeres. Maize knobs can promote the preferential inheritance (or meiotic drive) of particular chromosomes. The two different knob-specific maize satellites differ in their capacity to promote chromosome segregation [11], and several trans-acting factors that affect knob activation have been identified [39]. At least four distinct genetic functions mapping to Abnormal chromosome 10 (Ab10) play a role in the segregation of knobs; in some cases, these have been characterized through screens for the suppression of meiotic drive [39].

Two other models can account for the appearance of centromere functions in non-centromeric DNA. The first, centromere spreading, has been described in Drosophila. In this model, the migration of centromere functions to adjacent DNA is mediated by the deposition of centromere-binding factors (Figure 2b; [40]). Alternatively, centromeric and pericentromeric DNA can be recruited to repair the ends of broken chromosomes (Figure 2c). Support for this possibility is based on the discovery of large tracts of pericentromeric duplicated DNA (termed duplicons) that are thought to arise during the repair process. For example, mouse chromosomes 5 and 6 appear to be chromosome fission products, repaired with fragments of centromeric and pericentromeric DNA from other chromosomes [41]. Intriguingly, the duplications produced by these repair processes often appear to become hot-spots for further duplications and rearrangements.

Studies of centromere function

Current approaches for delineating the functional components of centromeres include: co-immunoprecipitating centromere proteins with their associated DNA fragments, measuring transmission rates of non-essential chromosome fragments with truncated centromeres, and assessing transmission rates of in-cyto-assembled artificial chromosomes. In Arabidopsis, antibodies to the CENP-A homolog co-immunoprecipitate 178 bp satellite repeats [42]; whereas in maize, CenH3 binds to both CentC and CRM [43]. These studies implicate satellites and retroelements as important DNA sequences for centromere function, but do not rule out a potential role for other DNA sequences in this function.

An important method for defining the functional domains of plant centromeres has relied on truncations of maize B chromosomes, which do not provide cellular functions and consequently can be lost without deleterious effects. Analysis of 25 B chromosomes that had truncated centromeres revealed a direct correlation between centromere size and meiotic transmission [13]. Centromere sizes of less than 1000 kb, or derivatives missing an internal 370 kb fragment, were lost at significantly higher rates. All of the derivatives were stable for multiple mitotic divisions, indicating that there may be genetic elements that distinguish meiosis and mitosis.

Assays that rely on artificial chromosome transmission would greatly facilitate the characterization of the functional components of plant centromeres. Such experiments have already been carried out using human artificial chromosomes: α-satellite arrays of 65 kb from chromosome 17 can confer mitotic centromere function, whereas chromosomes with similar arrays from the Y chromosome are lost during mitosis [44]. Interestingly, the Y chromosome arrays do not contain sites that bind to CENP-B, a centromere protein whose function is not completely understood. Future research using plant artificial chromosomes will enable analysis of the role of centromere length, satellite composition, and centromere-binding proteins during both meiosis and mitosis. Performing such experiments across different species will clarify the functional significance of the evolutionary changes that contribute to centromere diversity.

Conclusions

Plants hold great promise for clarifying the DNA elements that comprise centromere structure, determining the factors required for centromere function, and understanding the mechanisms that drive rapid centromere evolution. Limited comparisons of centromere DNA from species within the grass and crucifer families indicate that some DNA elements are conserved among closely related species. Further analysis of centromeres from many closely related plants will elucidate conserved functional elements. Plant artificial chromosomes will be useful for testing centromere function and identifying critical centromere elements. Functional assays and comparisons of centromere DNA sequence have far reaching implications, including improving our understanding of centromere evolution and its influence on speciation, allowing analysis of cis- and trans-acting factors that are important for proper inheritance of chromosomes, and permitting enhanced plant transformation using artificial chromosomes.

Acknowledgements

We thank members of the Preuss laboratory for helpful suggestions and critical reading of the manuscript. This work was supported in part by the National Science Foundation (grant to AEH and DP), a National Institute of Health National Research Service Award (to KCK), and the Howard Hughes Medical Institute.

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


6 Genome studies and molecular genetics


Using FISH analysis, the authors determine the centromeric distribution and organization of rice CentO satellites and CRR retroelements, including a quantification of the satellite arrays on all 12 chromosomes. Sequence alignments reveal homology between CentO and maize CentC, which is significant given the evolutionary separation of rice and maize. Finally, telomeric chromosomes derived from misalignments indicate that chromosome breakpoints map within CentO arrays, providing strong evidence that CentO is a functional component of rice centromeres.


CentO and a centromere-specific retrotransposon were used to identify two centromere BAC clones from a maize genomic library. Sequencing indicated that both clones consist of short CentC arrays interspersed with retrotransposons. A detailed description of the CentC and retrotransposon sequences is provided, which includes alignments and phylogenetic trees generated with centromere-specific elements from other grasses.


The authors of this paper describe the evolution of centromere satellite repeats in Arabidopsis thaliana ecotypes. Consensus sequences were derived for each of the ecotypes and the species, and the conserved and variable regions determined. The results presented here are indicative of the functionality of satellites and their rapid evolution even within a species.


This review describes recent advances in studies of satellite sequence evolution and satellite copy number. The authors discuss current hypotheses regarding satellite evolution and their relevance to speciation events.


The authors of this paper identify and analyze the functionality of satellites and their rapid evolution even within a species.


The authors use comparative FISH mapping in primates to document a centromere-repositioning event on chromosome 6. They propose this centromere-repositioning occurred because of an emerging species, rather than a transposition or inversion event, that relocated the endogenous centromere. They conclude that centromere emergence is
Evolution of plant centromeres

Hall et al.

612.0x794.0


This group demonstrates that antibodies to maize CENH3, a centromere-specific histone that marks functional centromeres, co-immunoprecipitate both CentC satellite and CRM retroelements. This work is significant because it suggests that, in addition to satellite DNA, a conserved retroelement may be a component of the functional maize centromere.


www.sciencedirect.com


actually fairly frequent in primates and, surprisingly, find that the repo-
positioning process does not affect neighboring chromosomal sequences.