Plant centromere organization: a dynamic structure with conserved functions

Jianxin Ma¹, Rod A. Wing², Jeffrey L. Bennetzen³ and Scott A. Jackson¹

Although the structural features of centromeres from most multicellular eukaryotes remain to be characterized, recent analyses of the complete sequences of two centromeric regions of rice, together with data from Arabidopsis thaliana and maize, have illuminated the considerable size variation and sequence divergence of plant centromeres. Despite the severe suppression of meiotic chromosomal exchange in centromeric and pericentromeric regions of rice, the centromere core shows high rates of unequal homologous recombination in the absence of chromosomal exchange, resulting in frequent and extensive DNA rearrangement. Not only is the sequence of centromeric tandem and non-tandem repeats highly variable but also the copy number, spacing, order and orientation, providing ample natural variation as the basis for selection of superior centromere performance. This review article focuses on the structural and evolutionary dynamics of plant centromere organization and the potential molecular mechanisms responsible for the rapid changes of centromeric components.

Introduction

The centromeres of eukaryotic chromosomes are responsible for sister chromatid cohesion (see Glossary) and for normal chromosomal segregation during mitosis and meiosis, which are essential for development and cellular proliferation in all organisms. These functions are conserved across species, but the DNA components that are involved in kinetochore formation differ greatly, even between closely related species [1–5]. Given the functional importance of centromeres and their rapidly diverging sequence, it is not surprising that many studies have focused on characterizing centromeres in terms of their structure, function and evolution. Deciphering the structural features and the sequence variation of centromeres is essential for a more complete understanding of centromere function(s). Here, we review recent studies on plant centromere organization with respect to the nature, timing, evolutionary processes and biological consequences of centromeric DNA amplification, recombination and rearrangement, thereby providing insights into the conserved functions of this dynamic structure.

Molecular components of plant centromeres

Repetitive DNA is ubiquitous and abundant in centromeric regions of higher eukaryotes. In all flowering plants investigated so far, different centromeres from an individual genome are generally composed of the same types of DNA component, mainly large arrays of centromeric satellite repeats and centromeric retrotransposons (CRs) [6–14]. However, the abundance and the arrangement of these repeats vary substantially, both within and among species [8,10,12,13,15]. In Arabidopsis thaliana, centromeres contain 180-base satellite repeats organized in tandem arrays that range from 0.4 to 1.4 Mb on different chromosomes. By contrast, the amount of CentO satellite DNA in rice (Oryza sativa) centromeres differs even more [10], ranging from 60 kb to 1.9 Mb on different chromosomes. Highly variable amounts of centromeric satellite DNA (CentC) were also detected in different maize (Zea mays) centromeres [6,14,15]. In general, these satellite repeats are specific to centromeric regions, with few copies and no long arrays found elsewhere in the genome [16–19], suggesting that these repeats are crucial for centromere function.

There is also substantial variation in the copy number of satellite repeats in homologous centromeric regions from different ecotypes [20], inbred lines [15] and subspecies.
[10]. For example, the centromeric region of *O. sativa* chromosome 6 contains over fourfold more copies of CentO in the *japonica* subspecies than in the *indica* subspecies [10]. A similar phenomenon was observed recently in different inbred lines of maize [15]. Given that the haplotypes compared in each of these studies diverged from each other between only a few thousand and half a million years ago [21], it is clear that rapid growth and/or shrinkage of centromeric satellite-repeat arrays can occur within short evolutionary time frames.

In addition to copy number variation among organisms, the sequence of centromeric satellite repeats differs markedly among organisms, even among closely related species [2–5]. Short conserved motifs are detected between CentO (rice) and CentC (maize) [10], suggesting that they originated from a common progenitor ~50–70 million years ago (Mya) [22]. However, a recent study using chromatin immunoprecipitation (ChIP) reported that CentO repeats were absent from the entire data set recovered from functional centromeres of *Oryza brachyantha*, a species of wild rice that last shared a common ancestor with rice only ~7–9 Mya [23,24]. A different centromeric satellite repeat, CentO-F, was found in all centromeres of *O. brachyantha*, with no detectable homologous copies in the entire *O. sativa* genome. This finding suggests that old centromeric satellite repeats have completely disappeared and that new centromeric satellite repeats have been generated in the short time since their independent descent from a common ancestor.

Centromere organization was also shown to vary greatly with respect to CRs, such as CRRs (in rice) [10] and CRM (in maize) [25]. Similar to other long terminal repeat (LTR) retrotransposon families in centromeric, pericentromeric and euchromatic regions [26], the CRRs with two intact LTRs that were identified in centromeric regions are very young evolutionarily, most having been inserted within the past few million years [11,27]. Therefore, it is not surprising that the distribution patterns and abundance of CRs among centromeres in an individual organism or among homologous centromeres from different species are highly variable [10,15,25,28]. In all (non-centromeric) orthologous regions of grass species that have been compared, including for rice, maize, wheat (*Triticum monococcum*), barley (*Hordeum vulgare*) and sorghum (*Sorghum bicolor*), few or no specific LTR-retrotransposon insertions are detected at orthologous loci in different species [29–34], largely owing to their rapid removal by illegitimate recombination and unequal homologous recombination [35,36]. However, some highly conserved CR sequences are present in the centromeric regions of most grasses that have been investigated [37,38]. This finding suggests that CRs evolve more slowly than other retroelements because they are selected for an important centromere function, or are located in a more slowly evolving portion of the genome, or both. An exception to the general conservation of grass CRs was recently reported in *O. brachyantha*, in which CRR-related sequences were found to be absent from all functional centromeres [23].

Despite their preferential accumulation in centromeric regions, CRRs are present in pericentromeric and euchromatic regions of all 12 rice chromosomes [28]. However, retrotransposons that are not CRs are also present in centromeric regions of maize [39] and rice [12,13]. It is particularly intriguing that ~50% of the retrotransposon sequences isolated by ChIP cloning in *Oryza rhizomatis*, a wild rice species that last shared a common ancestor with rice more recently than *O. brachyantha*, show clear homology to previously identified non-CRR families present in both centromeric and non-centromeric regions of rice [23]. This result supports the idea that non-CRRs could be converted to fill a functional role similar to CRRs. It is also possible that some non-CR sequences might be required for the maintenance of centromere function [25,40].

**Chromatin structures of plant centromeres**

Although there is tremendous divergence in centromeric DNA sequences and marked variation in centromere organization, centromeric chromatin have similar structural features in eukaryotes. In general, centromeric chromatin is distinguished from the surrounding pericentromeric heterochromatin by the presence of a specialized histone H3 called CENH3 (also known as CENP-A). CENH3 replaces the canonical histone H3 and interacts with other core histone proteins to form a specific type of nucleosome that is essential for kinetochore formation [1,41]. Species-specific CENH3 molecules have been identified in all eukaryotes investigated so far, including humans (*Homo sapiens*) [42], budding yeast (*Saccharomyces cerevisiae*) [43], *Drosophila melanogaster* [44], *Arabidopsis* [40], rice [11], maize [25], sugarcane (*Saccharum officinarum*) [45] and *Luizula nivea* [46]. Because CENH3 is present exclusively in functional domains of centromeres [1,4,47], ChIP-based analyses with a CENH3-specific antibody have been used routinely to characterize functional DNA components found to bind to CENH3, such as CentO and CRRs in rice [11] and CentC and CRM in maize [25]. In a comprehensive ChIP-based analysis using a set of tiled DNA sequences from rice *Cen8* (the centromeric region of chromosome 8), Nagaki *et al.* identified a ~750 kb CENH3-binding domain, which defines the boundaries of functional *Cen8* [11]. The interactions between CENH3 and centromeric repeats (both CRs and centromeric satellite repeats) also have been documented recently in other plants [45,46]. Therefore, it is improbable that this interaction is a universal phenomenon in the plant kingdom.

**Epigenetic features of plant centromeres**

Not all centromeric repeats in a centromeric region are associated with CENH3 [25,40]. For example, two ~1.4 kb CentO blocks (CentO-IV and CentO-V), which are separated from the other three main CentO blocks (CentO-I, CentO-II and CentO-III), are not found in the CENH3-binding domain of rice *Cen8* [11] (Figure 1). It is particularly intriguing that recent observations obtained from neocentromeres in primates [41] and barley [48] convincingly showed that the canonical centromeric repeats are dispensable for attaining some centromere functions. If normal centromeres are lost or inactivated, regions without centromeric repeats can recruit CENH3 and other centromere-associated proteins to assemble a
functional kinetochore that promotes chromosome segregation [48–51]. Therefore, it is probable that epigenetic mechanisms have essential roles in mediating centromere assembly, whereas, in most higher eukaryotes, centromeric repeats seem to accumulate after a new centromere has been formed [1,2], and the repeats provide a favorable environment for establishing centromeric chromatin, probably by recruiting sequence-specific binding proteins, to ensure the stable inheritance of centromeres [47,50].

An evolutionary model involving ‘centromere drive’ [2,52] has been proposed recently to explain the rapid sequence divergence of the genes encoding CENH3 proteins and centromere DNA repeats. According to this model, during female meiosis in plants and animals, centromeres compete by microtubule attachments for inclusion in the single meiotic product that becomes the egg nucleus and is preferentially transmitted to the next generation. This preferential inheritance, although beneficial to the next-generation hopes of a selfish centromere, would be detrimental to the equal likelihood inheritance of parental chromosomes in mendelian genetics. Therefore, it is expected that CENH3 proteins would be selected for nonpreferential properties at the same time that centromere cis components (i.e. DNA and epigenetically inherited chromatin features) would be selected for preferential segregation properties. This hypothesis is supported by the findings that CENH3 and/or CENP-C, a poorly conserved centromeric protein, have undergone positive selection, thereby suppressing centromere drive by restoring meiotic parity epigenetically in plants and animals [40,53]. By contrast, both CENH3 and CENP-C show signs of negative selection in budding yeast, in which centromere drive is predicted to be absent, given the consistent equal transmission of all parental centromeres [2]. This model also is reminiscent of the meiotic drive process for maize knob repeats (which are extra-centromeric satellites), which condition preferential segregation and transmission to progeny through female meiosis [54]. The rates of transmission of these knob repeats in female meiosis correlate with the sizes of satellite-repeat arrays [55]. Assuming that centromere variants with expanded satellite-repeat arrays increase microtubule-binding ability during female meiosis in plants and animals, preferential accumulation of centromeric satellite repeats during centromere evolution would be an expected outcome of centromere drive.

Rearrangement of centromeric sequences in Cen8 and Cen4 of rice
Segmental duplication and inversion of centromeric DNA revealed by structural and phylogenetic analyses
Recent in-depth analyses of the rice Cen8 and Cen4 sequences [27,56] have provided insights into the evolutionary dynamics, processes and molecular mechanisms that have resulted in rapid amplification and variation of centromeric DNA in higher eukaryotes. One unexpected observation is the presence of large, tandemly triplicated...
segments (96 kb, 90 kb and 26 kb, respectively) in the Cen8 region (Figure 1), accounting for 212 kb of the ~750 kb CENH3-binding domain. According to the insertion dates of shared LTR retrotransposons and the sequence identities between these three segments, it was estimated that two duplication events were involved and occurred successively at similar times within the last ~0.3 million years, and these were followed by partial deletion of one segment. In addition, on the basis of phylogenetic analysis of all satellite monomers in Cen8, 48 satellite monomers in CentO block I (in Cen8) were found to have respective best (98–99% identical) matches (i.e. monomers) arranged in opposite orientation but conserved order in CentO block III, indicating a recent segmental duplication of the CentO satellite-repeat arrays in the functional domain of Cen8.

A similar study was carried out on the core region of rice Cen4 [56]. Phylogenetic analysis of all 460 satellite monomers in this core region revealed several apparent segmental duplications of CentO satellite arrays and CRRs interspersed in the satellite arrays [56]. Hence, segmental duplication is likely to be a common process driving centromeric DNA amplification, probably in all plant centromeres. An unexpected observation was that most CRRs accumulated in the core region of Cen4 by rounds of segmental duplication rather than by integration of active elements. Given that most CRRs and the flanking CentO monomers were duplicated <0.3 Mya, it is now easier to understand why the O. sativa subspecies indica and japonica, which diverged from a common ancestor ~0.44 Mya [21], show high levels of haplotype variation in orthologous centromeric regions [10,27]. However, marked amplification of centromeric repeats does not result only from recent segmental duplications. More ancient duplications that have been deleted, have diverged or have been obscured by numerous more recent and overlapping DNA rearrangement events might no longer be detectable. Even in the most recently duplicated CentO satellite-repeat arrays (described earlier), insertions or deletions (indels) from a few monomers up to 130 monomers were frequent [27,56].

Because of the near-complete suppression of homologous chromosome exchange by recombination that is expected in all centromeric regions of rice [57], recombinational conversion and unequal homologous sister-chromatid recombination [58,59] are likely to be the main mechanisms underlying the duplication of CentO satellite-repeat arrays and the array-mediated duplication of CRRs [27,56]. The rates of these types of recombination might be increased by the abundance of highly homologous CentO satellite repeats. Although duplication of CentO satellite-repeat arrays and CRRs is frequent in the Cen4 core region, it should be noted that unequal recombination could also lead to deletion of centromeric repeats, as is predicted to occur in the process of solo-LTR formation. Given the active expansion of CentO satellite-repeat arrays in rice centromeres [10], there are probably selective forces, such as centromere drive [2,52], counteracting the loss of centromeric satellite DNA and CRRs by unequal recombination.

Inversions of some duplicated CentO satellite-repeat arrays were revealed in Cen8 and Cen4 (Figure 1). Unequal recombination between homologous sequences that are present in inverted orientation and interspersed in the CentO satellite-repeat arrays would cause inversion of the sequences between the recombining sites. Because of the abundance of CRRs arranged in different transcriptional orientations in the Cen4 core region, it is possible that recombination between CRRs has mediated the inversion of CentO arrays. This probable mechanism is supported by the presence of LTR retrotransposons (intact or fragmented) at five of the six junction sites of CentO blocks arranged in inverted orientation in Cen8 and Cen4 (Figure 1).

**Rapid removal of centromeric DNA by unequal recombination and illegitimate recombination**

Although LTR retrotransposons have preferentially accumulated in the Cen8 region and the Cen4 core region, processes that remove retrotransposon DNA in both regions are also active. These are reflected by the presence of solo-LTRs and various internally deleted or truncated elements [27,56]. Of the 245 LTR-retrotransposon elements or fragments identified in a 1.97 Mb region that contained Cen8, only 26.5% are intact elements, whereas solo-LTRs and partially deleted elements account for 24.9% and 46.2%, respectively [27]. Most solo-LTRs are the products of unequal homologous recombination between the two LTRs of a single element [35], whereas deleted or truncated elements have been generated mainly by illegitimate recombination, a mechanism that does not require extensive sequence homology [26,35]. Hence, unequal recombination and illegitimate recombination have removed most of the LTR-retrotransposon sequences from Cen8, and similar results were obtained for Cen4 [56].

Deletion of LTR-retrotransposon DNA by unequal recombination and illegitimate recombination in rice has been found to be exceptionally efficient. It was estimated that at least 190 Mb of retrotransposon DNA has been removed from the rice genome by these processes in the past 4 million years [21,36]. Because the intact LTR retrotransposons in the centromeric regions of rice are as young as those in non-centromeric regions, and because the relative percentages of intact and truncated elements are similar to those observed in the entire rice genome [36], it seems that the elimination of retrotransposon DNA is similarly efficient in all parts of the rice genome. This result agrees with similar findings in D. melanogaster [60], for which it has been concluded that rates of DNA sequence loss are identical in euchromatic and heterochromatic regions.

Perhaps the most interesting observation regarding the structural variation of LTR retrotransposons in Cen8 was the discovery of preferential accumulation of solo-LTRs in a subregion that includes the main CentO blocks in the CENH3-binding domain [27]. The ratio of solo-LTRs to intact elements in this subregion is approximately threefold more than in other centromeric and pericentromeric subregions and is slightly greater than in non-centromeric regions of rice [21], indicating a hot spot for unequal intra-element recombination in Cen8. Given the near-complete suppression of meiotic chromosomal exchange in centromeric regions that is indicated by recombinational mapping [57], and given the high ratio...
Box 1. Unanswered questions

Roles of centromeric sequences
Both CRs and centromeric satellite repeats interact with CENH3 in all plant species that have been investigated. Do these two classes of centromeric repeat have similar or distinct roles in centromere function? Non-CRs and single-copy sequences can also interact with CENH3 in some organisms [11,25,40]. Are these CENH3-binding sequences all required for centromere formation? The development of artificial plant chromosomes might provide new tools to study centromere structure and function. What are the minimal requirements for assembly of a highly efficient artificial chromosome?

Divergence of centromeric components
Centromere drive has been proposed to explain the rapid divergence of centromeric DNA and proteins (e.g., CENH3 and/or CENP-C) in most plants and animals. This model might also account for the disappearance of CentO and CRRs from functional centromeres of *Oryza brachyantha*, a wild rice species that diverged from *Oryza sativa* (rice) ~7–9 Mya. However, conserved CR sequences, centromeric satellite repeats and CENH3 were found in multiple grass species: for example, rice and maize, which diverged from each other ~50–70 Mya. Why have centromeric repeats undergone such marked changes in *O. brachyantha* but not in rice and maize? Are there genetic factors that could suppress the process of centromere drive?

Evolutionary stage of rice *Cen8*
The presence of active genes in the CENH3-binding domain of *Cen8* and of low quantities of CentO repeats in the *Cen8* region suggested that *Cen8* might represent an intermediate stage in the evolution of centromeres from genic regions, as in human neocentromeres, to fully mature centromeres that accumulate megabases of homogeneous satellite arrays [11]. However, rapid contraction or expansion, reshuffling and rearrangements of centromeric DNA are proposed to have occurred in *Cen8*. Do these changes disguise the evolutionary stage of *Cen8*?

Concluding remarks
In-depth analyses of plant centromeric sequences, including the first completely sequenced rice centromeres, have deepened our understanding of the dynamic structures, variation and evolution of higher eukaryotic centromeres. High rates of unequal (and presumably equal) recombination in the centromere have been detected, suggesting that crossover suppression is the main factor underlying the traditional perception of centromeres as recombinational cold spots. Rates of DNA removal from centromeric regions also do not seem to be slower than in other parts of the genome, suggesting that selection against inserts in euchromatic regions and/or preferential insertion into heterochromatin are the main reasons for uneven LTR-retrotransposon accumulation. The results indicate that centromere function can be maintained even with high rates of local genome rearrangement, and they suggest that this rearrangement could be a considerable component of centromere drive [2,44]. Therefore, in plants and animals, this provides a basis for competition for more successful female gamete transmission of a centromere that shows superior association with the kinetochore. Although insights into the dynamic structures of plant centromeres have been gained from these exciting observations, many intriguing questions (Box 1) are also raised and need to be further investigated.

Acknowledgements
We are grateful to the anonymous reviewers for their valuable comments and to the National Science Foundation for supporting this work (grant number DBI-0501814 to J.L.B., DBI-0227414 to S.A.J. and DBI-0321678 to R.A.W. and S.A.J.).

References
32 Ma, J. et al. (2005) DNA rearrangement in orthologous org regions of the maize, rice and sorghum genomes. Genetics 170, 1209–1220
39 Nagaki, K. et al. (2003) Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics 163, 759–770
43 Meluh, P.B. et al. (1998) Cse4p is a component of the core centromere of Saccharomyces cerevisiae. Cell 94, 607–613
46 Nagaki, K. et al. (2005) Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant Luzula nivea. Plant Cell 17, 1886–1893

AGORA initiative provides free agriculture journals to developing countries

The Health Internetwork Access to Research Initiative (HINARI) of the WHO has launched a new community scheme with the UN Food and Agriculture Organization.

As part of this enterprise, Elsevier has given hundreds of journals to Access to Global Online Research in Agriculture (AGORA). More than 100 institutions are now registered for the scheme, which aims to provide developing countries with free access to vital research that will ultimately help increase crop yields and encourage agricultural self-sufficiency.

According to the Africa University in Zimbabwe, AGORA has been welcomed by both students and staff. "It has brought a wealth of information to our fingertips", says Vimbai Hungwe. "The information made available goes a long way in helping the learning, teaching and research activities within the University. Given the economic hardships we are going through, it couldn’t have come at a better time."

For more information, visit www.aginternetwork.org