Chromosome Research (2007) 15:77–84
DOI: 10.1007/s10577-006-1104-z

Rice as a model for centromere and heterochromatin research

Huihuang Yan & Jiming Jiang*
Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA;
Fax: +1-608-262-4743; E-mail: jjiang1@wisc.edu
**Correspondence*

Key words: centromere, chromosome structure, heterochromatin, *Oryza sativa*, rice, transcription

Abstract

Rice (*Oryza sativa*) has become an important model plant species in numerous research projects involving genome, molecular and evolutionary biology. In this review we describe the reasons why rice provides an excellent model system for centromere and heterochromatin research. In most multicellular eukaryotes, centromeres and heterochromatic domains contain long arrays of repetitive DNA elements that are recalcitrant to DNA sequencing. In contrast, three rice centromeres and the majority of the cytologically defined heterochromatin in the rice genome have been sequenced to high quality, providing an unparalleled resource compared to other model multicellular eukaryotes. Most importantly, active genes have been discovered in the functional domains of several rice centromeres. The centromeric genes and sequence resources provide an unprecedented opportunity to study function and evolution of centromeres and centromere-associated genes.

Introduction

Rice is one of the most important model plant species for genome and molecular biology research. The rice genome has been sequenced several times using different sequencing approaches (Goff *et al*. 2002, Yu *et al*. 2002, Matsumoto *et al*. 2005), resulting in one of the best-sequenced genomes among all multicellular eukaryotes. Compared to other grass species rice can be readily transformed. Large populations of T-DNA and retrotransposon-based mutants have been developed (Hirochika 2001, Jeong *et al*. 2002, Sallaud *et al*. 2004). This combination of sequence and genetic resources provides an invaluable foundation for genome mapping and gene discovery in this important crop species. In this review we discuss using rice as a model for chromosome research, with a focus on centromeres and heterochromatin.

Sequencing and mapping of rice centromeres

The centromere is the most characteristic landmark on monocentric eukaryotic chromosomes, appearing as a constriction on condensed metaphase chromosomes. A large protein complex, the kinetochore, is assembled on the centromere to mediate the attachment of the microtubules. Centromeres are also enriched with cohesin and play a key role in maintaining proper sister chromatid cohesion. In multicellular eukaryotes, centromeres are often embedded within cytologically distinctive heterochromatin and are associated with megabase-sized arrays of highly homogenized satellite DNA (Henikoff *et al*. 2001, Jiang *et al*. 2003). This highly repetitive nature makes centromeres difficult for sequencing and fine-scale genetic mapping. The genomes of several model eukaryotes, including *Drosophila melanogaster*, human, mouse and *Arabidopsis thaliana*, have been

sequenced. However, none of the centromeres in these species has been fully sequenced.

Rice centromeres contain a 155-bp satellite repeat CentO (Dong *et al.* 1998, Nonomura & Kurata 2001, Cheng *et al.* 2002) and a centromere-specific retrotransposon CRR (Miller *et al.* 1998, Cheng *et al.* 2002). The CRR retrotransposons are highly enriched in the centromeres and are intermingled with the CentO satellite (Cheng *et al.* 2002, Nagaki *et al.* 2005). The total amount of CentO in any one centromere varies from ~60 kb to ~2 Mb (Cheng *et al.* 2002). Currently available sequencing technologies are overwhelmed by megabase-sized satellite DNA arrays, but it has proved possible to complete the sequence of the centromeres of rice chromosomes 4, 5 and 8 which contain less than 150 kb of CentO repeat (Nagaki *et al.* 2004, Wu *et al.* 2004, Zhang *et al.* 2004, Matsumoto *et al.* 2005). In addition, the rice chromosome pseudomolecules (http://www.tigr.org/tdb/e2k1/osa1/) extend into the CentO domains of eight of the remaining nine centromeres. The sizes of the CentO-containing gaps on these remaining nine centromeres were estimated to be ~150 kb to ~1.9 Mb based on fluorescence *in-situ* hybridization (FISH) data (Cheng *et al.* 2002). The sizes of the core domains of rice centromeres, which are marked by the centromeric histone 3 variant, CENH3, are ~1–2 Mb (Nagaki *et al.* 2004, Yan *et al.* 2006). Thus, at least portions of the core domains of some of the remaining nine rice centromeres are already covered by the current pseudomolecules. Taken together, rice provides the most extensive amount of centromeric sequences among all multicellullar eukaryotes.

The positions of all 12 centromeres on the rice genetic map were initially inferred by assigning restriction fragment length polymorphism (RFLP) markers to the respective arms using rice secondary and telotrisomics (Singh *et al.* 1996). These mapping positions were refined on the high-density rice genetic linkage map (Harushima *et al.* 1998) which contains over 3000 RFLP markers (http://rgp.dna.affrc.go.jp/E/Publicdata.html). Centromeres characteristically suppress genetic recombination, leading to co-segregation of multiple markers located within the recombination-suppressed domain. Thus, the mapping positions of the rice centromeres can be established by reconciling the locations of co-segregating RFLP markers and the CentO repeats on the rice pseudomolecules. The sizes of the recombination-suppressed domains of rice chromosomes 3 and 8 are 3113 kb and 2312 kb, respectively (Yan *et al.* 2005, 2006). The CENH3-associated core domains of these two centromeres are embedded within the recombination-suppressed domains (Yan *et al.* 2005, 2006) (Figure 1).

Transcription and histone modifications associated with rice centromeres

Traditionally, the centromere has been thought of as a highly heterochromatic and transcriptionally silent chromosomal domain. Discovery of active genes in the centromere of rice chromosome 8 (*Cen8*) showed that this concept must be revised (Nagaki *et al.* 2004). Rice *Cen8* contains a ~750-kb core domain associated with CENH3. There are at least 16 active genes within this domain (Yan *et al.* 2005). Similarly, the 1881-kb CENH3 domain of the centromere of rice chromosome 3 (*Cen3*) contains 19 genes for which evidence indicates transcription (Yan *et al.* 2006) (Figure 1). Mapping of 17-bp rice mRNA signatures and small RNA signatures, generated by massively parallel signature sequencing (MPSS), on the *Cen3* pseudomolecule revealed a highly complex transcriptome for this centromere (Yan *et al.* 2006). More than 40% of the mRNA signatures represent sequences transcribed from intergenic regions, introns or the antisense strand of exons. Numerous matches between *Cen3* sequence and small RNA signatures indicate that at least some of the repetitive DNA elements in rice centromeres are also transcribed (Yan *et al.* 2006). Active genes and their normal transcription have also been demonstrated in a human neocentromere (Saffery *et al.* 2003).

The extensive transcription in rice *Cen8* and *Cen3* is consistent with several recent reports on centromere-associated histone modifications in model eukaryotes. In the fission yeast *Schizosaccharomyces pombe*, the centromere cores are enriched with methylated histone H3 at Lys4 (H3K4me), which defines euchromatin, but not with methylated histone H3 at Lys9 (H3K9me), which marks heterochromatin (Cam *et al.* 2005). In contrast, the chromatin domains immediately flanking the centromere cores represent classical heterochromatin and are enriched with H3K9me and Swi6, a homologue of *D. melanogaster* Heterochromatin Protein 1 (HP1) (Cam *et al.* 2005). In *Drosophila* and humans, centromeric chromatins marked by CID/CENP-A (the respective CENH3 homologues) are distinct from both euchromatin and the flanking

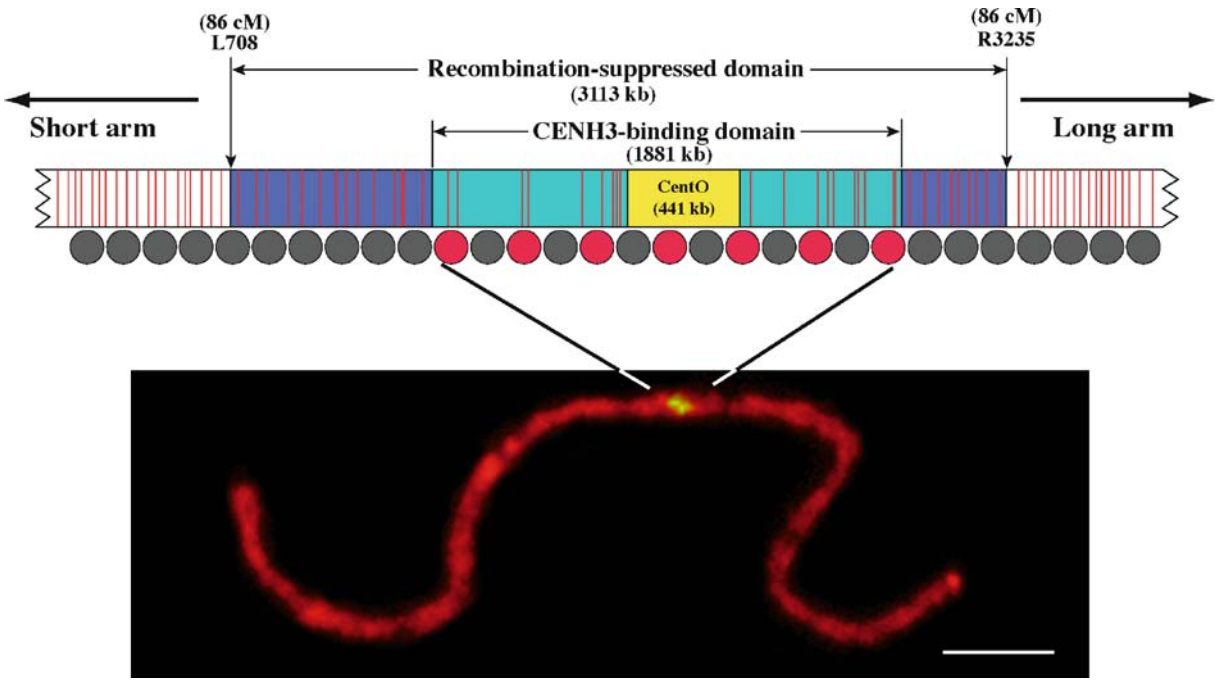

Figure 1. Structure of the centromere of rice chromosome 3. *Cen3* includes a ~1881-kb region associated with CENH3 (turquoise colour). This CENH3-binding domain includes ~441 kb of the CentO satellite repeat (yellow colour) and is embedded within a 3113-kb recombination-suppressed domain (blue colour). The boundary of the recombination-suppressed domain is marked by DNA markers L708 and R3235, both at 86 cM. The CENH3-binding domain contains active genes (red bars), but with a lower density than the flanking domains. Red solid circles and grey solid circles represent CENH3-associated and canonical H3-associated nucleosomes, respectively. The scale bar represents 5 μm. The chromosome image is courtesy of Zhukuan Cheng.

heterochromatin (Sullivan & Karpen 2004). Specifically, the CID/CENP-A-associated chromatin is hypoacetylated, and enriched with H3K4me2 but depleted of H3K4me3, H3K9me2, and H3K9me3 (Sullivan & Karpen 2004). H3K4 methylation and marker gene expression have also been demonstrated in human artificial chromosomes (Grimes *et al.* 2004, Lam *et al.* 2006). Wong *et al.* (2006) recently investigated the methylation status of CpG islands across a 6.76-Mb chromosomal region spanning a human neocentromere. The transcription within the neocentromere is associated with selective hypomethylation of pockets of sequences without compromising the overall silent chromatin state and function of the centromere (Wong *et al.* 2006). All these results show that centromeric chromatin is different from classical heterochromatin and is permissible for transcription.

Rice provides an excellent model to study transcription and histone modification associated with centromeres because the wealth of single and low copy sequences identified within centromeric domains can be used for specific DNA sequence-based analyses such as chromatin immunoprecipitation (ChIP) and microarrays. We recently measured the levels of H4 acetylation and H3K4/H3K9 methylation at 176 genes within a 3.5-Mb region spanning *Cen8*. Active genes in both the centromere core and the flanking regions showed enrichment of H4 acetylation and H3K4me2. We also detected pockets of H3K9me2 in *Cen8* (Yan *et al.* 2005). These results suggest the presence of both euchromatin and heterochromatin in rice centromeres. Thus, the CENH3-containing 'centrochromatin' does not appear to be distinguished by a unique combination of H3 and H4 modifications. In-depth studies on more histone modification patterns using microarray-based platforms will reveal characteristics of rice centromeric chromatin and differences, if any, between animal and plant centromeres.

Rice as a model for centromere evolution research

Although the centromeres have highly conserved and specific functions, the DNA sequences associated with the centromeres represent the most dynamic and mutable portion of the genome. Centromeric DNA sequences have no discernible conservation among different model organisms and can be significantly diverged among closely related species (Lee *et al.* 2005). The position of a centromere may relocate during evolution. A centromere can also be activated or inactivated without changing its associated DNA sequences. These facts make evolutionary study of centromeres a challenging endeavour.

Ventura *et al.* (2001) provided the first detailed analysis of 'centromere repositioning', in the mammalian X chromosome. The X chromosomes of black lemur, ringtailed lemur and humans share a highly conserved genetic synteny. However, the centromeres of the three X chromosomes are not only located in different positions but also contain different satellite repeats (Ventura *et al.* 2001), suggesting that emergence of new centromeres is the most likely explanation of the observed repositioning. Similar centromere repositioning has since been demonstrated for several other mammalian chromosomes (Eder *et al.* 2003, Ventura *et al.* 2004, Carbone *et al.* 2006). One possible mechanism of centromere repositioning is via neocentromere activation coupled with the loss or inactivation of the original centromere. In humans, non-α satellite-based neocentromeres and the α satellite-based but inactivated original centromeres can coexist and be stably maintained through both mitosis and meiosis (Tyler-Smith *et al.* 1999, Amor *et al.* 2004). Most importantly, although neocentromere formation results in substantial remodelling of chromatin associated with the core centromeric domain, it has no measurable effect on transcriptional competency of the underlying genes (Saffery *et al.* 2003). Thus, a neocentromere originating in a genic region is not detrimental to the function of its underlying genes. Both neocentromere formation and centromere inactivation have been reported recently in plant species (Nasuda *et al.* 2005, Han *et al.* 2006).

The structures of rice *Cen8* and *Cen3* as described above resemble intermediates between human neocentromeres, which are often associated with typical genic regions, and satellite repeat-based centromeres. Thus, *Cen8* and *Cen3* appear to represent relatively recent births of centromeres. During the neocentromere-mediated evolutionary process (Figure 2), satellite repeats from other centromeres may invade the emergent neocentromeres. Alternatively, new satellite repeats may emerge in the neocentromeres (Figure 2). A short starting satellite repeat array, such as the CentO array in rice *Cen8*, may expand via mechanisms that govern tandem repeat evolution across the genome (Charlesworth *et al.* 1986). Amplification of arrays within centromeric domains may be further accelerated by the action of meiotic drive (Malik & Henikoff 2002) in which the asymmetric nature of female meiosis provides an opportunity for competition between chromatids having different kinetochore activities (reviewed in Malik & Bayes 2006). The discovery that *Cen3* and *Cen8* in a wild rice species, *Oryza punctata*, contain 1.47 and 1.44 Mb arrays of the CentO repeats which almost perfectly overlay with CENH3 (Zhang *et al.* 2005), compared to ~440 kb and ~60 kb arrays in the equivalent cultivated rice domains, may provide a useful test of this model.

Rice has more than 20 wild relatives, including species closely related to cultivated rice and those

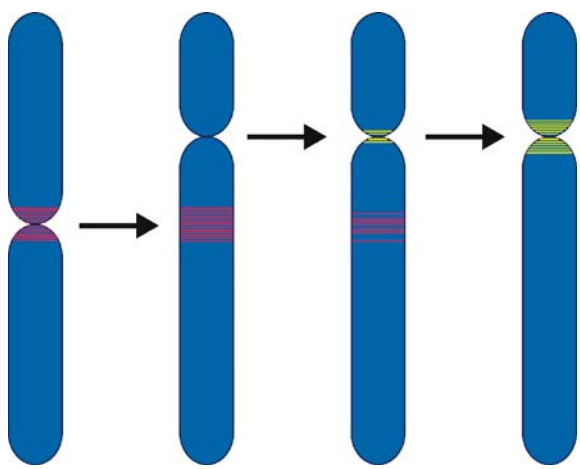

Figure 2. A model of neocentromere-mediated centromere evolution. The original centromere contains mainly satellite repeats (red line box). A neocentromere emerges, coupled with the inactivation of the original centromere via unknown epigenetic mechanisms. A new satellite repeat (yellow lines) emerges in the neocentromere. The satellite array in the inactivated centromere may degrade and eventually be eliminated during evolution. The new satellite repeat array expands and the survived neocentromere evolves to eventually become a typical satellite repeat-based mature centromere. The diagram is modified from Yan *et al.* (2006).

diverged from rice as much as ~9 million years (Khush 1997, Guo & Ge 2005). Several wild rice species do not contain the CentO repeat (Hass *et al*. 2003), suggesting that in these species the CentO repeats have either diverged significantly or been replaced by unrelated sequences. Thus, the wild rice species provide a system to study evolution of centromere-associated repetitive DNA sequences. We recently isolated several new centromeric satellite repeats from two CentO-less *Oryza* species, *O. rhizomatis* and *O. brachyantha*, and demonstrated different evolutionary patterns of centromeric DNA in these two species (Lee *et al*. 2005). A major genome project was launched with a goal to develop bacterial artificial chromosome (BAC)-based physical map of 11 wild *Oryza* species containing distinct genomes (http://www.omap.org/). The resource generated by this project will provide an opportunity to sequence *Cen8* in all these species and to reveal the evolution of a centromere among a group of closely related species. Comparative analysis of the orthologous regions of *Cen8* from these species will allow the establishment of patterns of centromeric DNA rearrangements and the dynamics of transposable element insertion and satellite amplification. Such analyses will also reveal how active genes in a centromere can adapt to a potentially heterochromatic and recombination-suppressed centromeric domain.

Heterochromatin distribution in the rice genome

Heterochromatin was originally described by Emil Heitz (1928) as relatively inaccessible chromatin that remains condensed (darkly stained) throughout the cell cycle. Traditionally a heterochromatic domain is defined cytologically, such as the heterochromatic 'knobs' described in maize (McClintock 1929, Brown 1949) and *Arabidopsis* (Fransz *et al*. 2000). Such heterochromatic features can be differentially stained on interphase nuclei or on meiotic pachytene chromosomes. Some large and highly condensed heterochromatic features are visible even on mitotic metaphase chromosomes. However, the definition of heterochromatin has now expanded to the submicroscopic and molecular level. A heterochromatic domain can be 'cryptic' and cytologically invisible but is associated with heterochromatic marks such as H3K9me and HP1. Such a 'cryptic' domain can be as small as a single gene.

The heterochromatic and euchromatic regions in rice are well differentiated on pachytene chromosomes (Khush *et al*. 1984, Cheng *et al*. 2001) (Figure 3). The majority of brightly 4′,6-diamidino-2-phenylindole (DAPI)-stained heterochromatin is distributed in the proximal regions of rice chromosomes (Cheng *et al*. 2001). Notably, on chromosomes 4 and 10, heterochromatin covers almost the entire short arms and the proximal portions of their long arms, with the remaining regions being largely euchromatic (Figure 3). In addition to these major heterochromatic domains, smaller, knob-like heterochromatic domains are also visible at the distal ends of several chromosomes. The heterochromatin distribution pattern is highly conserved between the *indica* and *japonica* subspecies (Cheng *et al*. 2001). This conservation has also been observed in several wild rice species (Zhukuan Cheng, personal communication). Most importantly, the majority of the heterochromatic domains in rice have been tiled by the existing physical map and sequenced to high quality. The current rice sequence maps contain over 50 Mb finished sequences from the major heterochromatic regions (http://www.tigr.org/tdb/e2k1/osa1/). The sequences can be anchored to specific heterochromatic domains by FISH. The deep coverage of rice heterochromatin by both physical and sequence maps provides the prerequisite for a comprehensive characterization of its sequence composition, transcriptional regulation, DNA and histone modifications.

Heterochromatin has been well defined cytologically in several other plant species, including tomato (Khush & Rick 1968), *A. thaliana* (Fransz *et al*. 1998), sorghum (Kim *et al*. 2005) and maize (Shi & Dawe 2006, Wang *et al*. 2006). However, cytologically defined heterochromatin is almost exclusively located in the centromeric regions in *A. thaliana* (Fransz *et al*. 1998) while the vast majority of the *A. thaliana* genes are located within euchromatin. High-quality genome sequences are not available in other plant species. Thus, rice provides the best available system to study heterochromatin organization and its consequences for gene expression.

Transcription and epigenetic modifications of rice heterochromatin

Heterochromatin was traditionally viewed as a repository of 'junk' DNA and is well known for its silencing

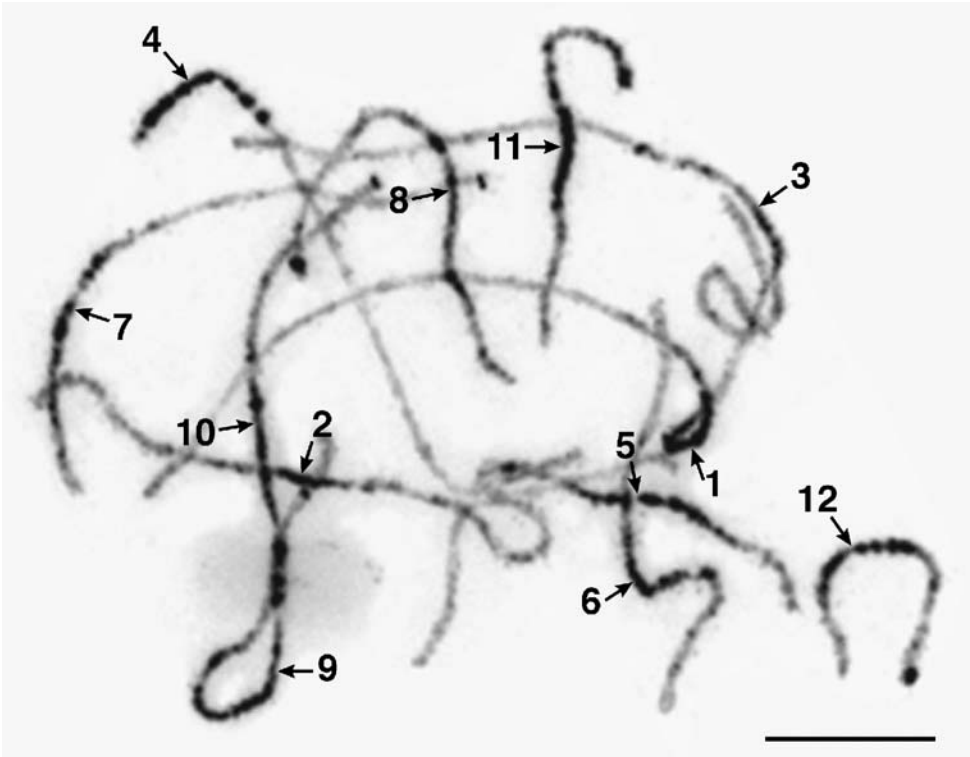

Figure 3. Differential staining of euchromatic and heterochromatic regions on pachytene chromosomes of rice (*indica* rice variety Zhongxian 3037). Chromosomes were stained with DAPI and were converted as black-and-white to enhance the contrast of the image. Arrows point to the centromeres that were identified by FISH mapping of the CentO satellite repeat on the same cell (data not shown). The scale bar represents 10 μm. Figure reproduced with permission from Cold Spring Harbor Laboratory Press © 2001, from Cheng *et al.* (2001).

effect on gene expression (Dimitri *et al.* 2004). However, transcribed genes have been identified from heterochromatic regions in several model organisms (Dimitri *et al.* 2004, Yasuhara & Wakimoto 2006). In *D. melanogaster*, dozens of genes with critical function in viability or fertility have been identified from the centric heterochromatin (Hoskins *et al.* 2002, Yasuhara & Wakimoto 2006). These genes are tolerant of the silencing effect of heterochromatin and, in some cases, depend on heterochromatin proteins for their normal expression (Yasuhara *et al.* 2005). In plants a ~730-kb heterochromatic knob on the short arm of *Arabidopsis* chromosome 4 (hk4S) has been fully sequenced and extensively characterized (Fransz *et al.* 2000, McCombie *et al.* 2000). This heterochromatic knob contains ~30 known or hypothetical genes plus significant numbers of transposable elements and related repeats that are associated with both DNA methylation and H3K9me (Lippman *et al.* 2004). Most of the genes showed normal expression and appeared to reside in the 'euchromatic islands' that were not associated with DNA methylation and H3K9me, thus differing from the 'true' heterochromatic genes described in *D. melanogaster* that can be expressed only in a heterochromatic environment (Yasuhara & Wakimoto 2006).

The cytologically defined heterochromatic regions in rice are associated with extensive transcription, including single-copy protein-coding genes, based on matches with rice EST/fl-cDNA sequences (http://www.tigr.org/tdb/e2k1/osa1/) and expression profiling using genomic tiling microarrays (Jiao *et al.* 2005, Li *et al.* 2005, 2006). The heterochromatin on rice chromosome 4 is most pronounced (Figure 3), covering nearly half (~17 Mb) of the entire chromosome (35.2 Mb). This heterochromatic half of the chromosome has a suppressed genetic recombination, being 2.3 cM/Mb, versus 4.8 cM/Mb for the euchromatin half (Zhao *et al.* 2002). The heterochromatic half contains significantly more repeats, except

miniature inverted-repeats transposable elements that are known to be preferentially associated with genes (Feng *et al*. 2002). As expected, the annotation of chromosome 4 revealed that the heterochromatic region showed ~3-fold enrichment of transposable element-related genes (65.3 genes/Mb vs 22.7 genes/Mb) but ~4-fold decrease of active genes (17 genes/Mb vs 65.5 genes/Mb), as compared with the euchromatic region (http://www.tigr.org/tdb/e2k1/osa1/).

Cytological data indicated that the ~17 Mb heterochromatin region of rice chromosome 4 was organized as eight intensively stained domains, usually one to several Mb in size, and weakly stained intervening domains on pachytene chromosome (Cheng *et al*. 2001, Jiao *et al*. 2005) (Figure 3). Consistent with the cytological pattern, these intensively stained domains showed decreased levels of transcription as compared with their flanking domains (Jiao *et al*. 2005). Research on epigenetic modifications of rice heterochromatin is still in its infancy. The DNA methylation and histone modification patterns are largely unknown, as well as what impact, if any, such patterns have on gene expression and regulation. However, the availability of the cytological data, genome sequence and recently developed techniques, such as ChIP-chip, will allow such questions to be addressed in the future.

Acknowledgements

We thank Tim Langdon for his valuable comments on the manuscript. Rice centromere research was supported by grant FG02-01ER15266 from DOE and grant 2006-35604-16649 from USDA-CSREES to J.J.

References

Amor DJ, Bentley K, Ryan J *et al*. (2004) Human centromere repositioning 'in progress'. *Proc Natl Acad Sci USA* **101**: 6542–6547.

Brown WL (1949) Numbers and distribution of chromosome knobs in United States maize. *Genetics* **34**: 524–536.

Cam HP, Sugiyama T, Chen ES, Chen X, FitzGerald PC, Grewal SIS (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nat Genet* **37**: 809–819.

Carbone L, Nergadze SG, Magnani E *et al*. (2006) Evolutionary movement of centromeres in horse, donkey, and zebra. *Genomics* **87**: 777–782.

Charlesworth B, Langley CH, Stephan W (1986) The evolution of restricted recombination and the accumulation of repeated DNA sequences. *Genetics* **112**: 947–962.

Cheng Z, Buell CR, Wing RA, Gu M, Jiang J (2001) Toward a cytological characterization of the rice genome. *Genome Res* **11**: 2133–2141.

Cheng ZK, Dong F, Langdon T *et al*. (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *Plant Cell* **14**: 1691–1704.

Dimitri P, Corradini N, Rossi F, Verni F (2004) The paradox of functional heterochromatin. *Bioessays* **27**: 29–41.

Dong F, Miller JT, Jackson SA, Wang GL, Ronald PC, Jiang J (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc Natl Acad Sci USA* **95**: 8135–8140.

Eder V, Ventura M, Ianigro M, Teti M, Rocchi M, Archidiacono N (2003) Chromosome 6 phylogeny in primates and centromere repositioning. *Mol Biol Evol* **20**: 1506–1512.

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

Fransz P, Armstrong S, Alonso-Blanco C, Fischer TC, Torres-Ruiz RA, Jones GH (1998) Cytogenetics for the model system *Arabidopsis thaliana*. *Plant J* **13**: 867–876.

Fransz PF, Armstrong S, de Jong JH *et al*. (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: Structural organization of heterochromatic knob and centromere region. *Cell* **100**: 367–376.

Goff SA, Ricke D, Lan TH *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). *Science* **296**: 92–100.

Grimes BR, Babcock J, Rudd MK, Chadwick B, Willard HF (2004) Assembly and characterization of heterochromatin and euchromatin on human artificial chromosomes. *Genome Biol* **5**: R89.

Guo YL, Ge S (2005) Molecular phylogeny of Oryzeae (Poaceae) based on DNA sequences from chloroplast, mitochondrial, and nuclear genomes. *Am J Bot* **92**: 1548–1558.

Han FP, Lamb JC, Birchler JA (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc Natl Acad Sci USA* **103**: 3238–3243.

Harushima Y, Yano M, Shomura A *et al*. (1998) A high-density rice genetic linkage map with 2275 markers using a single F_2 population. *Genetics* **148**: 479–494.

Hass BL, Pires JC, Porter R, Phillips RL, Jackson SA (2003) Comparative genetics at the gene and chromosome levels between rice (*Oryza sativa*) and wildrice (*Zizania palustris*). *Theor Appl Genet* **107**: 773–782.

Heitz E (1928) Das heterochromatin der Moose. *Jahrb Wiss Botanik* **69**: 762–818.

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

Hirochika H (2001) Contribution of the *Tos17* retrotransposon to rice functional genomics. *Curr Opin Plant Biol* **4**: 118–122.

Hoskins RA, Smith CD, Carlson JW *et al*. (2002) Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biol* **3**: research0085.

Jeong DH, An SY, Kang HG *et al*. (2002) T-DNA insertional mutagenesis for activation tagging in rice. *Plant Physiol* **130**: 1636–1644.

Jiang J, Birchler JB, Parrott WA, Dawe RK (2003) A molecular view of plant centromeres. *Trends Plant Sci* **8**: 570–575.

Jiao YL, Jia PX, Wang XF *et al*. (2005) A tiling microarray expression analysis of rice chromosome 4 suggests a chromosome-level regulation of transcription. *Plant Cell* **17**: 1641–1657.

Khush GS (1997) Origin, dispersal, cultivation and variation of rice. *Plant Mol Biol* **35**: 25–34.

Khush GS, Rick CM (1968) Cytogenetic analysis of the tomato genome by means of induced deficiencies. *Chromosoma* **23**: 452–484.

Khush GS, Singh RJ, Sur SC, Librojo AL (1984) Primary trisomics of rice: origin, morphology, cytology and use in linkage mapping. *Genetics* **107**: 141–163.

Kim JS, Islam-Faridi MN, Klein PE *et al*. (2005) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. *Genetics* **171**: 1963–1976.

Lam AL, Boivin CD, Bonney CF, Rudd MK, Sullivan BA (2006) Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc Natl Acad Sci USA* **103**: 4186–4191.

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

Li L, Wang X, Xia M *et al*. (2005) Tiling microarray analysis of rice chromosome 10 to identify the transcriptome and relate its expression to chromosomal architecture. *Genome Biol* **6**: R52.

Li L, Wang XF, Stolc V *et al*. (2006) Genome-wide transcription analyses in rice using tiling microarrays. *Nat Genet* **38**: 124–129.

Lippman Z, Gendrel AV, Black M *et al*. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 471–476.

Malik HS, Bayes JJ (2006) Genetic conflicts during meiosis and the evolutionary origins of centromere complexity. *Biochem Soc Trans* **34**: 569–573.

Malik HS, Henikoff S (2002) Conflict begets complexity: the evolution of centromeres. *Curr Opin Genet Dev* **12**: 711–718.

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

McClintock B (1929) Chromosome morphology in *Zea mays*. *Science* **69**: 629.

McCombie WR, de la Bastide M, Habermann K *et al*. (2000) The complete sequence of a heterochromatic island from a higher eukaryote. *Cell* **100**: 377–386.

Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics* **150**: 1615–1623.

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

Nagaki K, Neumann P, Zhang D *et al*. (2005) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. *Mol Biol Evol* **22**: 845–855.

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

Nonomura K, Kurata N (2001) The centromere composition of multiple repetitive sequences on rice chromosome 5. *Chromosoma* **110**: 284–291.

Saffery R, Sumer H, Hassan S *et al*. (2003) Transcription within a functional human centromere. *Mol Cell* **12**: 509–516.

Sallaud C, Gay C, Larmande P *et al*. (2004) High throughput T-DNA insertion mutagenesis in rice: a first step towards in silico reverse genetics. *Plant J* **39**: 450–464.

Shi J, Dawe RK (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics* **173**: 1571–1583.

Singh K, Ishii T, Parco A, Huang N, Brar DS, Khush GS (1996) Centromere mapping and orientation of the molecular linkage map of rice (*Oryza sativa* L.). *Proc Natl Acad Sci USA* **93**: 6163–6168.

Sullivan BA, Karpen GH (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol* **11**: 1076–1083.

Tyler-Smith C, Gimelli G, Giglio S *et al*. (1999) Transmission of a fully functional human neocentromere through three generations. *Am J Hum Genet* **64**: 1440–1444.

Ventura M, Archidiacono N, Rocchi M (2001) Centromere emergence in evolution. *Genome Res* **11**: 595–599.

Ventura M, Weigl S, Carbone L *et al*. (2004) Recurrent sites for new centromere seeding. *Genome Res* **14**: 1696–1703.

Wang C-JR, Harper L, Cande ZW (2006) High-resolution single-copy gene fluorescence in situ hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. *Plant Cell* **18**: 529–544.

Wong NC, Wong LH, Quach JM *et al*. (2006) Permissive transcriptional activity at the centromere through pockets of DNA hypomethylation. *PLoS Genet* **2**: e17.

Wu JZ, Yamagata H, Hayashi-Tsugane M *et al*. (2004) Composition and structure of the centromeric region of rice chromosome 8. *Plant Cell* **16**: 967–976.

Yan HH, Jin WW, Nagaki K *et al*. (2005) Transcription and histone modifications in the recombination-free region spanning a rice centromere. *Plant Cell* **17**: 3227–3238.

Yan HH, Ito H, Nobuta K *et al*. (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. *Plant Cell* **18**: 2123–2133.

Yasuhara JC, Wakimoto BT (2006) Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet* **22**: 330–338.

Yasuhara JC, DeCrease CH, Wakimoto BT (2005) Evolution of heterochromatic genes of *Drosophila*. *Proc Natl Acad Sci USA* **102**: 10958–10963.

Yu J, Hu SN, Wang J *et al*. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *indica*). *Science* **296**: 79–92.

Zhang W, Yi C, Bao W *et al*. (2005) The transcribed 165-bp CentO satellite is the major functional centromeric element in the wild rice species *Oryza punctata*. *Plant Physiol*. **138:** 1205–1215.

Zhang Y, Huang YC, Zhang L *et al*. (2004) Structural features of the rice chromosome 4 centromere. *Nucleic Acids Res* **32**: 2023–2030.

Zhao Q, Zhang Y, Cheng ZK *et al*. (2002) A fine physical map of the rice chromosome 4. *Genome Res* **12**: 817–823.