Mol Gen Genomics (2006) 275: 421–430
DOI 10.1007/s00438-006-0103-2

ORIGINAL PAPER

**Weidong Bao · Wenli Zhang · Qiuying Yang
Yu Zhang · Bin Han · Minghong Gu
Yongbiao Xue · Zhukuan Cheng**

Diversity of centromeric repeats in two closely related wild rice species, *Oryza officinalis* and *Oryzarhizomatis*

Received: 15 September 2005 / Accepted: 12 January 2006 / Published online: 4 February 2006

Abstract *Oryza officinalis* (CC, $2n = 24$) and *Oryza rhizomatis* (CC, $2n = 24$) belong to the *Oryza* genus, which contains more than 20 identified wild rice species. Although much has been known about the molecular composition and organization of centromeres in *Oryza sativa*, relatively little is known of its wild relatives. In the present study, we isolated and characterized a 126-bp centromeric satellite (CentO-C) from three bacterial artificial chromosomes of *O. officinalis*. In addition to CentO-C, low abundance of CentO satellites is also present in *O. officinalis*. In order to determine the chromosomal locations and distributions of CentO-C (126-bp), CentO (155 bp) and TrsC (366 bp) satellite within *O. officinalis*, fluorescence in situ hybridization examination was done on pachytene or metaphase I chromosomes. We found that only ten centromeres (excluding centromere 7 and 2) contain CentO-C arrays in *O. officinalis*, while centromere 7 comprises CentO satellites, and centromere 2 is devoid of any detectable satellites. For TrsC satellites, it was detected at multiple subtelomeric regions in *O. officinalis*, however, in *O. rhizomatis*, TrsC sequences were detected both in the four centromeric regions (*CEN* 3, 4, 10, 11) and the multiple subtelomeric regions. Therefore, these data reveal the evolutionary diversification pattern of centromere DNA within/or between close related species, and could provide an insight into the dynamic evolutionary processes of rice centromere.

Keywords Centromere · *Oryza* · CC genome

Communicated by W. R. McCombie

W. Bao · W. Zhang · Q. Yang · Y. Xue · Z. Cheng (✉)
State Key Laboratory of Plant Genomics,
National Center for Plant Gene Research,
Institute of Genetics and Developmental Biology,
Chinese Academy of Sciences, 100101 Beijing, P.R. China
E-mail: zkcheng@genetics.ac.cn
Fax: +86-10-64873428

Y. Zhang · B. Han
National Center for Gene Research,
Shanghai Institutes for Biological Sciences,
Chinese Academy of Sciences, 200233 Shanghai, P.R. China

M. Gu
Department of Agronomy, Yangzhou University,
225009 Yangzhou, P.R. China

W. Bao · W. Zhang · Q. Yang
Graduate School of the Chinese Academy of Sciences,
100049 Beijing, P.R. China

Introduction

Centromere is the specialized chromosome region where kinetochore complex is assembled to serve as the attachment site for spindle microtubules. Centromere plays an essential role in the correct transmission of chromosomes in both mitosis and meiosis. On the monocentric chromosome, the centromere is microscopically recognizable as the primary constriction.

Although kinetochore function is highly conserved throughout all eukaryotes, the underlying primary centromeric DNA sequences are considerably variable, even in closely related species (Henikoff et al. 2001; Kawabe and Nasuda 2005). In the simplest model of *Saccharomyces cerevisiae*, the functional centromere spans only around 125-bp long, and is termed as the "point centromere" (Clarke and Carbon 1985). However, in most of the eukaryotes, centromere DNA is typically very long and complicated, consisting of up to megabase-sized arrays of tandem repetitive sequences (Cleveland et al. 2003; Sun et al. 1997). To date, various centromeric repeats have been identified in a number of animals and plants, such as the most studied alpha-satellites (171-bp) in primates (Willard 1991; Alexandrov et al. 2001) and pAL1 satellite (180-bp) in *Arabidopsis thaliana* (Martinez-Zapater et al. 1986; Murata et al. 1994). However, there are also some plants, such as field bean and *Tradescantia*, where no centromere-specific tandem repeats could be detected (Houben and Brandes 1996). How centromeric satellites are involved in the activity of cen-

tromere, particularly at the evolutionary level, is poorly understood. To date, two different hypotheses have been proposed to explain the dynamic changes of centromeric tandem repeats: ''molecular drive'' hypothesis (Henikoff and Malik 2002) and ''library'' hypothesis (Ugarkovic and Plohl 2002). According to the former hypothesis, the rapidly evolving centromeric sequences could be explained by the genetic conflict and interaction between centromeric DNA and DNA-binding kinetochore proteins, such as centromere-specific histone H3 variant (CENH3) and CENP-C, which are shown under positive selection (adaptive evolution) in the respective species (Talbert et al. 2004), whereas the latter model depicts the different satellite sequences within a species as a set of satellites (satellite library), of which the copy number and sequence of individual satellite and the satellite profile vary between species, possibly due to various mechanisms of satellite dynamics.

As a good monocot model system, rice has been extensively studied for decades. Both molecular and cytogenetic characterizations about its centromere structure and organization have been well documented in recent years. Blocks of a 155-bp/or 165-bp satellite (CentO) monomers arranged in a tail-to-head pattern are mapped to the centromeres of all 12 chromosomes in *Oryza sativa*. These 155-bp/165-bp tandem repeats have been referred by different names in different research: RCS2 (Dong et al. 1998), TrsD (Kumekawa et al. 2001) and CentO (Cheng et al. 2002). In addition to CentO, the so-called centromeric retrotransposons (CRs), a member of Ty3/gypsy-type retrotransposon, was also found enriched preferentially at the centromeric regions. Sequences of CRs from rice (CRRs) are highly homologous with its counterpart in maize (CRM), and both CRR and CRM had been shown cytologically either intermingled with centromeric satellite or nested in the centromeric regions for megabase size (Jiang et al. 2003; Miller et al. 1998). Chromatin immunoprecipitation (ChIP) analyses using antibody against CENH3, homolog of CENP-A in human, has demonstrated that both CentO and CRRs are the functional components of centromere/kinetochore complex (Nagaki et al. 2004).

The genus *Oryza* has about 25 recognized species, which has been classified into seven diploid genome types (AA, BB, CC, DD, EE, FF and GG) and four allotetraploid genome types (BBCC, CCDD, HHKK and HHJJ), mainly based on the analysis of chromosome pairing behavior in interspecific hybrids and comparative genome in situ hybridization (Vaughan et al. 2003). Phylogenic study among different genome types has indicated that BB and CC genomes have a more close relationship with AA genomes than other genomes (Ge et al. 1999). Previous investigations have revealed that CC genome contains much less CentO sequences, but contains a number of CC genome-specific satellites, including a 366-bp TrsC sequence (also known as pOo2) (Nakajima et al. 1996; Zhao et al. 1989) and two centromere-associated satellites, CentO-C1 and CentO-C2, isolated by ChIP from *Oryza Rhizomatis* (Lee et al. 2005). In the present study, we report the isolation and characterization of a CentO-C1-like centromeric sequence, CentO-C, from *Oryza officinalis* by screening bacterial artificial chromosome (BAC) clones, and the chromosomal location of four repeated sequences, CentO, CentO-C, TrsC and CRR in both *O. officinalis* and *O. rhizomatis* ($2n = 2x = 24$, CC). The results manifested that both CentO-C and CentO satellites are centromere associated in *O. officinalis* and *O. rhizomatis*; however, TrsC satellite locations are more variable in *O. officinalis* and *O. rhizomatis*. Our observations in the two closely related rice species, therefore, could provide an insight into the dynamic evolution of centromere satellites and its potential influence on centromere function.

Materials and methods

Plant materials

Oryza officinalis (IRGC Acc. 101114) and *O. rhizaomatis* (IRGC Acc. 103421) were used for the following experiments. Genomic DNA was extracted from young leaves using the CTAB method (Murray and Thompson 1980). Two rice Nipponbare BAC clones, OSJNBb0088N06 and OSJNBa0048B17, were used as chromosome-specific fluorescence in situ hybridization (FISH) markers. Their genomic loci were obtained from the rice FPC website (http://www.genome.clemson.edu/projects/rice/fpc/) and Gramene (http://www.gramene.org/Oryza_sativa/).

Southern blot hybridization

Southern blots were prepared by digesting approximately 5 μg genomic DNA with selected restriction endonucleases, separating the digests on 1% agarose gel, and blotting them onto Hybond N+ membranes (Amersham) by capillary transfer. Hybridization probes were prepared by labeling the plasmid inserts fragment with [α-^{32}P]dCTP using the Prime-a-Gene® Labeling System (Promega, Madison, WI, USA) and hybridization overnight at 65°C. The washing conditions were: 2×, 1× and 0.5× SSC with 0.1% SDS, 15 min each at 65°C. The signals were recorded using the Storm scanner.

BAC library construction and screening

Bacterial artificial chromosome library was constructed for *O. officinalis* from nuclei preparations according to the protocol described by Texas A&M BAC Center (http://www.hbz.tamu.edu/bacindex.html). Endonuclease *Bam*HI partially digested genomic DNA was inserted into *Bam*HI site of plndigoBAC5 vector (Epicentre) and the ligation was used to transform *Escherichia coli* DH10B. A total of 23,040 clones were

stored in 384-well plates and 3,456 clones were transferred onto Hybond N+ filters (Amersham) for screening. Sheared *O. officinalis* genomic DNA was labeled with [α-^{32}P]dCTP using the Prime-a-Gene Labeling System (Promega) to screen for BAC clones, which contain a high amount of repetitive sequence according to the standard method (Sambrook and Russell 2001). Chromosomal locations of the candidate clones were checked by FISH.

DNA sequencing, assembly and data analyses

Bacterial artificial chromosome clones were sequenced by the shotgun method. Briefly, sheared BAC DNA fragments (2–3 kb) were cloned into PUC18 plasmid and sequenced by BigDye Terminator Cycle Sequencing V2.0 Ready Reaction (Applied Biosystems). The raw sequence data with 10× sequence coverage were assembled with PHRED and PHRAP program first, followed with careful manual check and refinement to minimize misalignments caused by the repeat sequence. The BAC sequences were submitted to the GenBank with accession numbers AY955100 (BAC 39O03), AY955098 (BAC 39G20) and AY955099 (BAC 39E03).

Homologous sequences were searched in the GenBank database using BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). The CRR subfamily classification is according to Nagaki et al. (2005). Tandem repeats in each BAC were identified and characterized by the software Tandem Repeats Finder (version 3.21) (Benson 1999) or dotplot with MegAlign™ (LaserGene). Multiple sequences alignments were first made using Clustal W (Thompson et al. 1997) and then adjusted manually and displayed using the BioEdit program (Hall 1999). The Sequence Logo was created with the online WebLogo server (Crooks et al. 2004). Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 3.1 (Kumar et al. 2004). Sequence evolutionary distance was calculated with the *p*-distance model. The phylogenetic tree was constructed by the neighbor-joining method with the Jukes–Cantor model as implemented by MEGA.

FISH probes and labeling

CRR (Centromeric retrotransposon in Rice) retrotransposon sequences, including pRCS1, pRCH1, pRCH2, pRCH3, pRCE1 and pRCE2, were derived from *O. sativa* and described previously by Dong et al. (1998). These six clones were combined to be used as CRR probes. CentO probe, CentO-C probe and TrsC probe were cloned from polymerase chain reaction (PCR) fragments of each satellite family. The primers derived from the conserved portions of each repeat family were used for PCR amplification (Table 1). The PCR products were cloned into T-easy vector (Promega) and the identity of each clone was confirmed by sequencing analysis. The PCR conditions were: 35 cycles at 94°C for 30 s, 55°C for 30 s and 72°C for 1 min. The probes were labeled with either digoxigenin-11-dUTP or biotin-dUTP (Roche) by standard nick translation (Dong et al. 1998).

Fluorescence in situ hybridization and immunostaining

Meiotic chromosomes were prepared as described by Cheng et al. (2001). The hybridization mixture consisted of 5 ng/μl DNA probes and 80 ng/μl of sheared salmon sperm DNA in 50% formamide, 10% dextran sulphate, 0.1% SDS and 2× SSC (0.3 M NaCl, 0.03 M sodium citrate). The probes and chromosomal DNA were hybridized at 37°C overnight. To investigate the relative chromosomal locations of two or more repetitive sequences, we hybridized the probes to pachytene spreads either simultaneously or sequentially. Digoxigenin-labeled probes were detected with fluorescein isothiocyanate-conjugated sheep-antidigoxigenin (Roche). Biotin-labeled probes were detected with Texas red-conjugated avidin (Vector Laboratories, Burlingame, CA). The chromosomes were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) in an antifade solution (Vector). Slides were visualized and recorded using a chilled CCD camera mounted on Olympus BX61 fluorescence microscope. Grayscale images were captured for each color channel and then merged by IPLab software.

For immunostaining, fresh panicle were harvested and fixed immediately in 4% paraformaldehyde for 20 min, anthers at the pachytene stage were picked and squashed. The rabbit antibodies to rice CENH3 (Nagaki et al. 2004) were diluted to 1:2,000 in TNB buffer containing 0.1 M Tris-HCl, pH 7.5, 0.15 M NaCl and 0.5% blocking reagent (Roche). 100 μl diluted antibodies was added to each slide, and the slides were incubated in a humid chamber at 37°C for 3 h. After three washes in PBS, 100 μl of rhodamine-conjugated goat anti-rabbit

Table 1 Primers used for amplification CentO, CentO-C and TrsC satellites

Satellite family	Primers sequences (5′-3′)	Reference sequences
CentO-F	CAA AA(A/C) TCA TGT TT(T/G) GGT G(A/T/G/C)	BX890594, AF058902
CentO-R	GGA CAT ATA G(G/T)A GTG (G/T)AT (A/G/T/C)	
CentO-C-F	CGT GTG CTA CGA AAC GAA (A/G/C/T)	AY955098, AY955099, AY955100
CentO-C-R	GCA ACC AAT GGT GCC AAA C	
TrsC-F	TGCCCTAAAACGCATCGCAT	AF286064–AF286070
TrsC-R	ATGTCCTTAAAGTTTCGA	

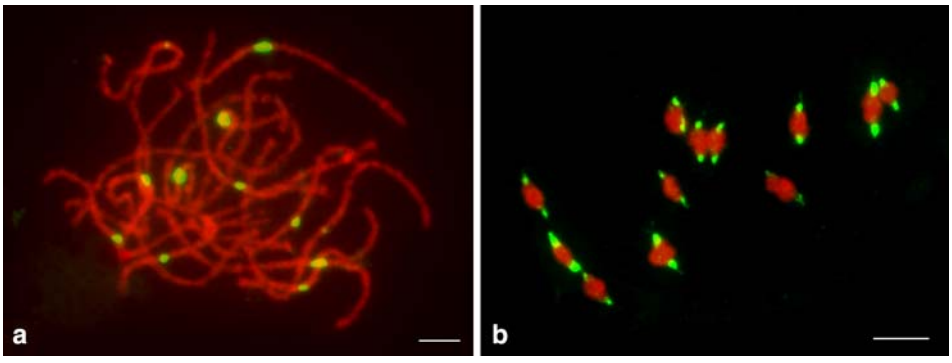

Fig. 1 Fluorescence in situ hybridization signals of centromeric bacterial artificial chromosome 39G20 (*green*) on *Oryza officinalis* pachytene chromosomes (**a**) and on metaphase I bivalent (**b**), chromosomes are pseudocolored as *red*. All bars, 5 μm

antibody solution (Vector) (1:50 in TNB buffer) was added to the slides. Incubation and washes were the same as for the primary antibody. The slides were counterstained with DAPI before being checked with a microscope. For FISH analyses after indirect immunofluorescence with CENH3 antibody, the slides were washed three times in 2×SSC at 42°C, hybridized with Digoxigenin-labeled probes and further detected with the proper antibody.

Results

Isolation of the centromeric sequences from *O. officinalis*

In order to isolate the centromeric repetitive sequence of *O. officinalis*, a BAC library consisting of 23,040 clones was constructed. We screened a total number of 3,456 clones with ^{32}P-labeled *O. officinalis* genomic DNA and got 19 clones showing strong hybridization signals. Each of these 19 BAC clones was labeled as FISH probes and hybridized to the pachytene chromosomes of *O. officinalis*. As a result, nine BACs were revealed with strong hybridization signals in the centromeric regions of all the chromosomes. A representative FISH image of these centromeric BACs on pachytene chromosomes or metaphase I chromosomes are shown by BAC 39G20 (Fig. 1).

To define the sequence composition of the centromere in *O. officinalis*, sequence analysis was done on three centromeric BAC clones, 39G20, 39O03 and 39E03. After 10× coverage shotgun sequencing and assembling, we got a total of ~66 kb insert sequences from these three BACs. The sequence size of each BAC is 6.879 kb for 39O03 (finished quality), 20.086 kb for 39G20 (four ordered contigs at phase II: 6.671, 7.446, 1.674 and 4.295 kb) and 38.148 kb for 39E03 (five ordered contigs at phase II: 12.270, 6.228, 7.898, 6.526 and 5.226 kb). The fingerprint gel shows that the sequence size accorded well with the physical size of each BAC (Fig. 2a). Dot plot analysis and BLASTN searching in GenBank revealed that the three BACs primarily contained a 126-bp satellite sequence and CRR-related sequences (CRR1 and noaCRR1) (Nagaki et al. 2005) and some unique-like sequences as well (Fig. 2b). The 126-bp satellite sequences accounted for around 35, 32 and 50% in 39O03, 39G20 and 39E03, respectively. In addition, we confirmed that the 126-bp satellite was also present in the remaining six centromeric BACs by PCR using primers designed from the conserved region of the satellite (data not shown), suggesting the prevalence of CentO-C in centromeric regions. We designate this 126-bp satellite as CentO-C.

Sequence distance analyses among CentO-C monomers showed that the average sequence distance within 39O03, 39G20 and 39E03 BAC is 0.152 ± 0.028, 0.192 ± 0.026 and 0.118 ± 0.018, respectively, which are a little bit smaller than the average distance between BACs: 0.186 ± 0.033 (39O03-39E03), 0.191 ± 0.023 (39G20-39E03) and 0.202 ± 0.028 (39O03-39G20). As CentO-C Logo consensus sequences show, the majority of CentO-C from these BACs can be distinguished at a number of nucleotide sites (Fig. 2c), indicating a locus or chromosome-specific homogenization process. In addition, we compared these CentO-C monomers with CentO-C1 monomeres (DQ058468–DQ058497), the recently isolated 126-bp centromere satellite from *O. rhizomatis* with ChIP method (Lee et al. 2005), and constructed a neighbor-joining phylogenetic tree. As a result, CentO-C monomers from individual BAC are separately clustered together, while CentO-C1 monomeres are scattered more widely (Fig. 2d), a finding that is readily understood from the fact that CentO-C1 monomers were derived from multiple centromeric loci (Lee et al. 2005). Figure 2d also reveals that CentO-C and CentO-C1 actually belong to the same sequence family, and it appears that CentO-C/CentO-C1 satellites had diverged very little after *O. officinalis* and *O. rhizomatis* split.

CentO, another centromeric satellite in *O. officinalis* and *O. rhizomatis*

Previous investigations have shown that the CentO repeat family is the major centromeric satellite in *O. sativa* and it occurs at all the 12 centromeres. However, CentO was not detected in *O. officinalis* and *O. rhizomatis* by Southern blot (Lee et al. 2005), but CentO/TrsD had been reported existing at a relatively low abundance (350 copies per haploid) in *O. officinalis* (Kumekawa et al.

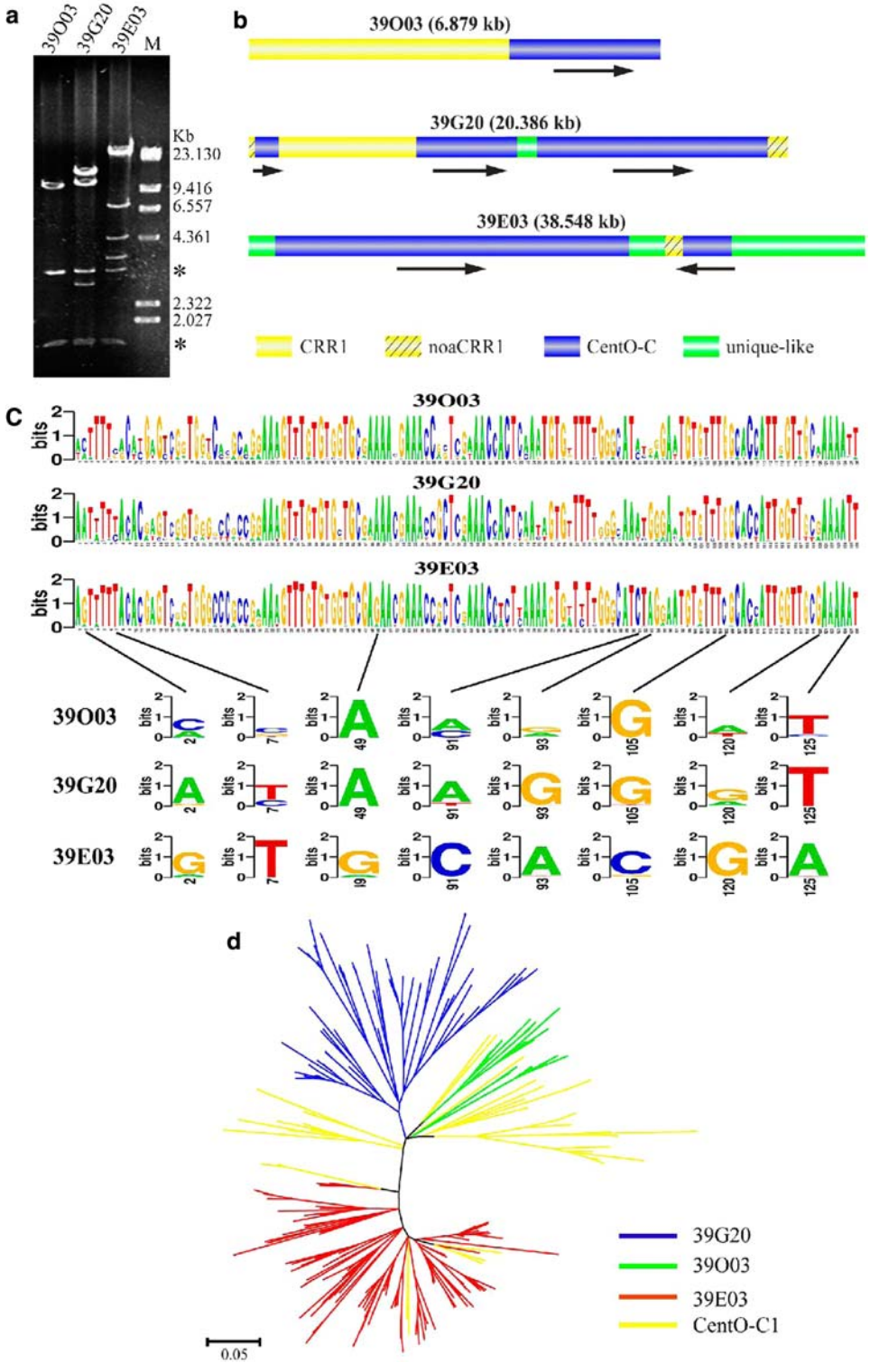

Fig. 2 Sequence analyses of bacterial artificial chromosomes (BACs) and CentO-C family. **a** Fingerprinting analyses of the BAC physical size. *Pst*I-digested BACs DNA are separated on 0.7% agarose gel. Note only partial vector (7.506-kb) is cut into two *Pst*I-bands, 1.541-kb and 3.017-kb (denote with *asterisk*), and the two remaining sessions (375-bp and 2.573-kb at each end) are fused with the *Pst*I-bands of insert. **b** Schematic sequence organization in BAC 39O03, 39G20 and 39E03. The relative orientations of the CentO-C arrays are shown by *black arrows* below. The size of each BAC sequence is shown in parenthesis. **c** Logos display the consensus sequences of CentO-C repeats (*upper panel*) and selected nucleotides significantly different between BACs (*lower panel*). **d** Bootstrap consensus tree of CentO-C monomers from the three BACs and the CentO-C1 monomers from *Oryza rhizomatis* (Lee et al. 2005). Neighbor-joining method, γ parameter of 2 for rate variation between sites, bootstrap value = 500 replication are options in the calculation. *Scale bar* represents estimated substitutions per site

2001). To test its presence in our selected accessions, Southern blot hybridization was performed. As most CentO satellite sequences derived from *O. officinalis* and *O. rhizomatis* are devoid of *Msp*I restriction site (which is present in *O. sativa*), but they contain *Hin*fI cutting site. In order to cut CentO into smaller size and show the ladder pattern of CentO, different restriction enzymes are used for different species. The existence of CentO in *O. officinalis* and *O. rhizomatis* genome was confirmed, although the latter one contains much less CentO (Fig. 3a). This result was also verified by PCR amplification with different PCR amplifications with different annealing temperatures (Fig. 3b).

To test these relatively small quantities of CentO sequences that are also located at the centromeric region in *O. officinalis*, we FISH-hybridized CentO probes against the pachytene chromosomes, of which the centromeric location was marked with anti-CENH3 antibody (Fig. 4a). As shown in Fig. 4c, only one CentO signal could be consistently detected overlapping with the CENH3 signal. Such distribution pattern of CentO sequences was further confirmed by repeated experiments, even under a lower hybridization stringency condition (30% formamide, data not shown). Therefore, CentO satellites are clustered at a single locus in the *O. officinalis* genome, and most likely located in the functional centromeric region of this particular centromere.

To further illustrate the relative locations of CentO and other centromeric sequences (CentO-C and CRR) in *O. officinalis*, CentO signals were washed off with 50% formamide in 2× SSC solution at 60°C for 5 min, and the same slide was sequentially reprobed with CentO-C and CRR. Strikingly, there were only ten centromeres that could be consistently turned on by CentO-C, excluding the CentO-containing centromere (Fig. 4b), whereas CRR could be hybridized to all the 12 centromeres (Fig. 4d). Thus, two CentO-C-minus centromeres are present in *O. officinalis*, one exclusively containing CentO and CRR sequences while the other one containing only CRR sequences (Fig. 4c, d). To determine which chromosome contains CentO sequences, we used the karyotype information of *O. officinalis* (data not shown), aided with chromosome-specific BAC FISH results, and revealed that it was chromosome 7 (Fig. 4i). Similarly, another CentO-C-minus centromere was identified to be the centromere of chromosome 2 (Fig. 4j). In addition, we estimated the size of CentO array of centromere 7 (*CEN7*) using the extended DNA fiber FISH technique. Eight independent CentO fiber signals (only six showed in Fig. 4k) were measured with

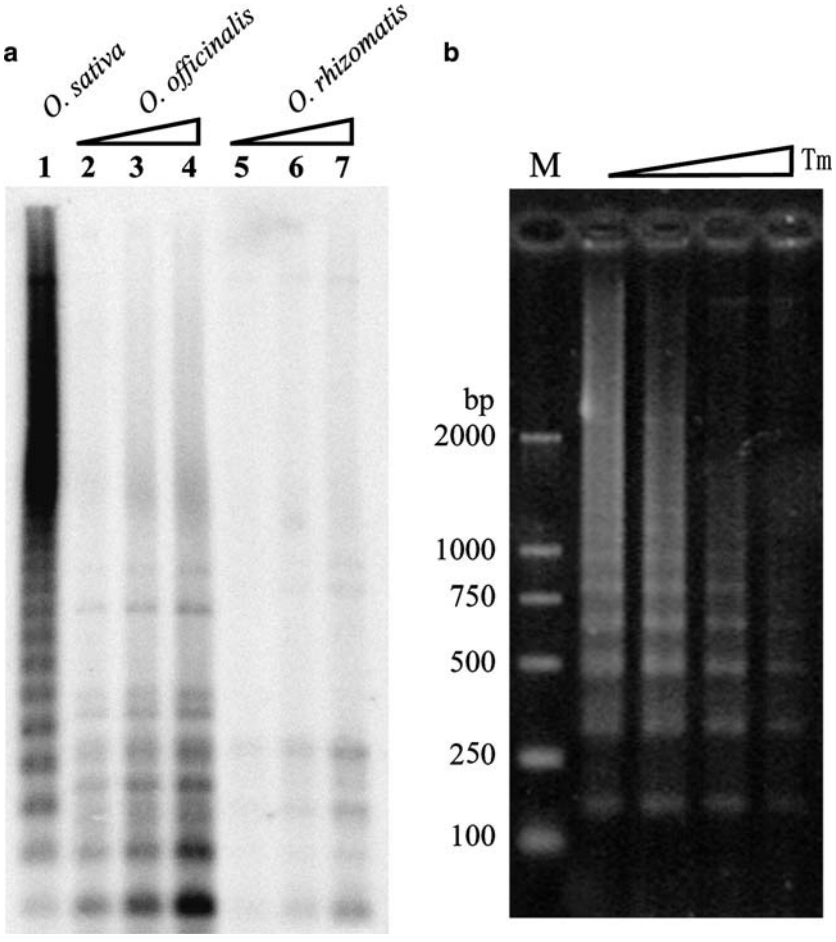

Fig. 3 Presence of CentO repeats in *Oryza officinalis* and *Oryza rhizomatis*. **a** Southern hybridization analysis of CentO to *Msp*I-digested genome DNA from *Oryza sativa*, *Hin*fI-digested genome DNA from *O. officinalis* and *O. rhizomatis*. *Lanes* 1, 2, 5 are loaded with approximate 5 μg DNA, while *lanes* 3, 6 and 5, 7 are loaded twofold and fourfold. **b** PCR amplification of CentO fragments from genomic DNA of *O. officinalis*, which shows a typical 155-bp ladder pattern of satellite repeat on agarose gel. PCR primers are showed in Table 1. Anneal temperatures (T_m) are increased from 50 to 60°C

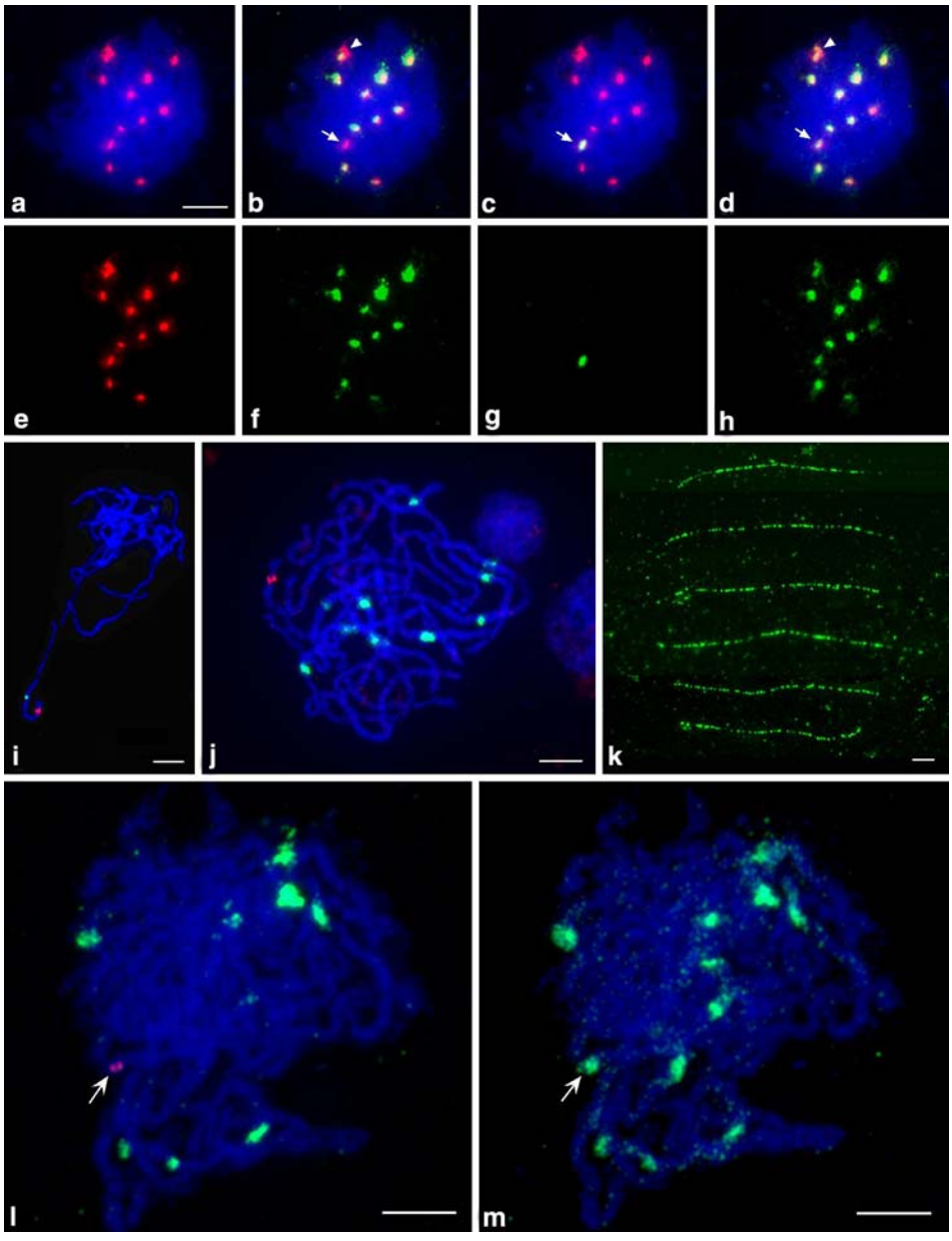

Fig. 4 Diversity of the centromeres in *Oryza officinalis* and *Oryza rhizomatis*. **a**–**d** Twelve centromere positions at pachytene stage are first immunostained with anti-centromere-specific histone H3 variant (CENH3) antibody (**a**, *red*), and then are sequentially detected with CentO (**c**), CentO-C (**b**) and CRR (**d**) probes, respectively. Only 10 out of 12 centromeres are consistently detected co-located with CentO-C probes (**a**, **b**), leaving two CentO-C-minus centromeres (*arrow* and *arrowhead*, **b**). One CentO-C-minus centromere co-located with the very single detectable CentO locus (**c**, *arrow*), and contained some CRR sequences as well (**d**, *arrow*), while the other CentO-C-minus centromere contains only CRR sequences (**d**, *arrowhead*). **e**–**h** Signals digitally separated from (**a**) to (**d**), CENH3 (**e**, *red*), CentO-C (**f**), CentO (**g**) and CRR (**h**), respectively. **i**, **j** Identifying the two chromosomes without CentO-C. The bacterial artificial chromosome (BAC) clone OSJNBa0048B17 (chromosome 7, *red*) are positioned on one of the two chromosomes that lack CentO-C signals but harbors CentO arrays (*green*) (**i**), while the other CentO-C-minus chromosome is identified by BAC clone OSJNBb0088N06 (chromosome 2, *red*) (**j**). **k** Six independent CentO signals on fiber fluorescence in situ hybridization (FISH) analysis. **l**, **m** FISH hybridization of CentO (**l**, *red*), CentO-C (**l**, *green*) and CRR (**m**, *green*) in *O. rhizomatis*. The centromere with CentO is arrowed in both **l** and **m**. All bars, 5 μm

an average length of 45.93 ± 4.26 μm, corresponding to 147.44 ± 13.67 kb according to a 3.21 kb/μm conversion rate (Cheng et al. 2002).

To examine the distribution of CentO in *O. rhizomatis*, FISH hybridization was conducted. As shown in Fig. 4l, a single weak CentO signal was also revealed to be overlapping with the CRR sequences (Fig. 4m), suggesting that CentO occurred at the centromere region. Again, this CentO locus is also avoided of any detectable CentO-C sequences, just like that in *O. officinalis*.

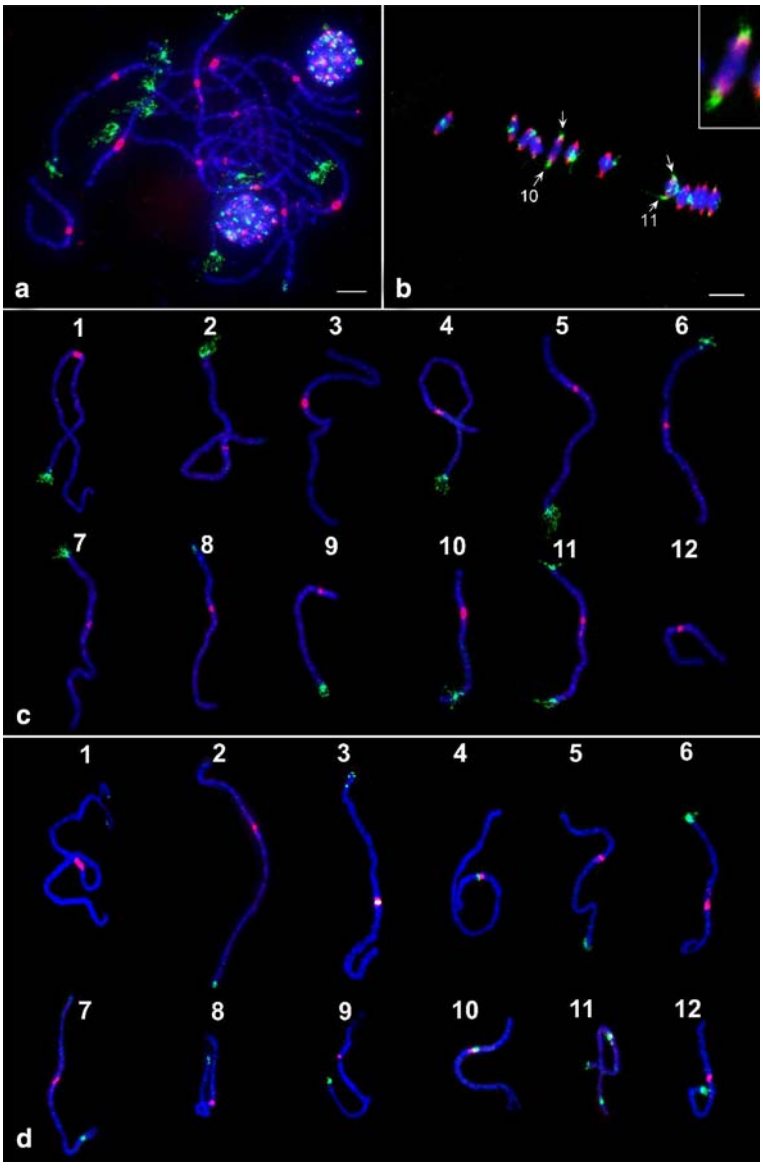

Fig. 5 The distributions of TrsC signals (*green*) in *Oryza officinalis* (**a**, **c**) and *Oryza rhizomatis* (**b**, **d**). Centromeres (*red*) are marked with CentO-C and CRR probes. TrsC (*green*) and centromere probes (*red*) are co-hybridized on the pachytene chromosomes of *O. officinalis* (**a**), or metaphase I chromosomes of *O. rhizomatis* (**b**), chromosomes 10 and 11 which are marked with *arrows* and *numbers*. *Inset* corresponds to the close-up of chromosome 10 (**b**). **c**, **d** Distributions of TrsC (*green*) on individual chromosomes in *O. officinalis* (**c**) and *O. rhizomatis* (**d**), note the unconserved TrsC positions between the two species and the centromeric TrsC signals on chromosomes 3, 4, 10 and 11 in *O. rhizomatis* (**d**). All bars, 5 μm

The subtelomeric repeats TrsC in *O. officinalis* are transferred to the centromere region in *O. rhizomatis*

TrsC is a previously identified satellite sequence with 366-bp unit length, and a high amount of TrsC was exclusively found in rice CC genome species (Nakajima et al. 1996; Zhao et al. 1989). Our FISH mapping from multiple images with TrsC probe revealed a total number of 11 subtelomeric TrsC signals in *O. officinalis*, distributed on ten chromosomes (Fig. 5a, c). However, 14 TrsC loci are situated not only at subtelomeric regions, but also occur at the centromeric regions of chromosomes 3, 4, 10 and 11 in *O. rhizomatis* (Fig. 5d). Moreover, the position of subtelomeric TrsC loci is not conserved between homologous chromosomes in the two species, such as chromosomes 4 and 10, etc. (Fig. 5c, d). Of those four TrsC-containing centromeres, centromere 3 (*CEN3*) and 4 (*CEN4*) comprise much less TrsC sequences according to the signal strength, while *CEN10* and *CEN11* contain much more TrsC sequences (Fig. 5d), which can be easily detected poleward on both chromosomes at metaphase I stage (Fig. 5b). Interestingly, although CentO-C and TrsC blocks of chromosome 10 [represented by CentO-C + CRR signals (red) and TrsC signals (green) in Fig. 5d] are closely juxtaposed (Fig. 5d), however, it is the TrsC arrays that are stretched most poleward at metaphase I stage, rather than the flanking CentO-C/CRR arrays (Fig. 5b, inset), which suggests that the major cluster of TrsC arrays substitute most, if not all, CentO-C arrays to be the functional component of *CEN10* kinetochore. Furthermore, we compared TrsC sequences (D85604) with CentO-C2 satellites (DQ058499–DQ058514), the recently isolated 366-bp centromeric repeat from *O. rhizomatis* by ChIP method (Lee et al. 2005), and we found that they share a high percentage of identity and actually belong to the same family (data not shown), supporting that TrsC arrays of *CEN10* are involved in the kinetochore function.

Discussion

Chromosome-specific centromere composition in *O. officinalis* and *O. rhizomatis*

Centromeres of most eukaryotes studied so far are composed of variable length of tandem repeats (Jiang et al. 2003; Kuznetsova et al. 2005). In this study, we reported identification of a 126-bp satellite, CentO-C, from *O. officinalis* by BAC screening and FISH. Compared with the 126-bp centromeric satellites (CentO-C1) isolated by ChIP cloning from *O. rhizomatis* (Lee et al. 2005), CentO-C monomers show a very high degree of similarity, but their species origins could be hardly distinguished by the sequence alone (Fig. 2d), which is readily accounted by their close evolutionary relationship between the two species. In addition to CentO-C satellites, we also confirmed by Southern blot the presence of low abundance of CentO sequences in both *O. officinalis* and *O. rhizomatis* (Fig. 3), although CentO sequences have been previously reported to be absent in CC genomes by Southern blot (Lee et al. 2005), which could be partly explained by the difference between different accession lines, or when different washing stringencies were applied. Most strikingly, the relatively low amount of CentO sequences is found to be exclusively concentrated at a single centromere in both *O. officinalis* and *O. rhizomatis* (Fig. 4j, l). In one accession of *O. officinalis* (Acc. 101114), CentO sequences occur at *CEN7* (Fig. 4j), and such a feature was also found for *CEN7* in another accession of *O. officinalis* (Acc. W1275, data not shown).

For the TrsC satellite, although it is distributed among subtelomeric loci in *O. officinalis* (Fig. 5a), it seems that TrsC have been translocated to four centromeres in *O. rhizomatis* (Fig. 5d). Thus, two (CentO, CentO-C) or three different centromeric satellites (CentO, CentO-C and TrsC) are present in *O. officinalis* and *O. rhizomatis*, respectively (Figs. 4, 5), contrasting with the situation in *O. sativa*, where CentO satellites are universal to all the 12 centromeres. In *Antirrhinum majus* (Zhang et al. 2005) and *Brassica* species (Harrison and Heslop-Harrison 1995), different centromeres are also reported comprising different satellites. Theoretically, the presence of chromosome-specific centromere satellites could be explained by the different ratios between the homogenization and mutation processes that happened to a particular satellite locus (Dover 2005; Ugarkovic and Plohl 2002). Thus, the diversity of the centromeric satellites found in *O. officinalis* or *O. rhizomatis* may only represent a snapshot showing the dynamic drift of the satellite populations. Nevertheless, the evolution of centromeric satellites have been made inevitably even more complicated by its association with various kinetochore proteins, such as CENP-B proteins (Kipling and Warburton 1997), CENP-A and CENP-C proteins (Amor et al. 2004). Evidence of positive selection (adaptive evolution) has been reported for CENP-A and CENP-C in some animals and plants, and was used to account for the rapid evolution and the complexity of centromeric DNA sequences (Henikoff et al. 2001; Talbert et al. 2004).

In the case of centromere 2 (*CEN2*) in *O. officinalis*, we consistently failed to detect the three satellites (CentO, CentO-C and TrsC). This could be due to the absence of detectable satellites by FISH, but it is still possible that *CEN2* contains other unidentified satellites. In plants such as field bean or *Tradescantia*, normal centromeres also appear to be devoid of specific satellites (Houben and Brandes 1996). Besides, the numerous neocentromeres identified in humans also harbor no significant repetitive sequences (Amor and Choo 2002). These data imply that centromeric satellites are not absolutely needed for centromere function.

TrsC sequences appear to be transposed to centromeric regions

TrsC satellite was first reported as rice CC genome specific (Zhao et al. 1989), but later investigations have revealed that TrsC repeats are not strictly limited to CC genomes and are also present at lower abundance (50–100 per haploid) in other rice species (Nakajima et al. 1996). In *O. officinalis*, all 11 TrsC loci are located at the subtelomeric regions (Fig. 5c), however, for the 14 TrsC loci in *O. rhizomatis*, 10 of them are subtelomeric and 4 are centromeric (Fig. 5d). Compared with the centromeric CentO or CentO-C sequences, the occurrence of TrsC at centromeric region in *O. rhizomatis* could be a relatively recent event. As shown by Lee et al. (2005), CentC, CentO and CentO-C are members of an ancient satellite family and they still share a low degree of sequence homology, thereby they all can act as the centromeric satellites in the respective species. In contrast, most TrsC loci of *O. officinalis* and *O. rhizomatis* are subtelomeric and are not conserved between homologous chromosomes (Fig. 5c, d), suggesting its capability to transpose over evolutionary time. In consistent with this presumption, telomeric simple repeats have been reported to occur at the centromeric regions in the *Solanum* species (Tek and Jiang 2004) and the *Pinus* species (Hizume et al. 2002), probably through a similar process. Furthermore, the *Tai*I satellite family in *Triticeae* species has been reported to be located at either the centromeric or the subtelomeric regions (Kishii and Tsujimoto 2002). Overall, our data imply that the four centromeric TrsC loci in *O. rhizomatis* were derived from the subtelomeric TrsC sequences, most likely through translocation and regional amplification. Intriguingly, we observed that the TrsC portions of *CEN10*, rather than the flanking CentO-C/CRR regions, were more likely the kinetochore assembling regions (Fig. 5b), which implies that the kinetochore assembling site is not be determined by the primary sequences of centromeric satellites alone, and other potential factors may be involved, such as CENH3 (Henikoff and Dalal 2005).

Acknowledgments The anti-OsCENH3 peptide antibody against rice CENH3 was kindly provided by Dr. S. Henikoff (Howard Hughes Medical Institute). This work was supported by grants from the Ministry of Sciences and Technology of China (2005CB120805), the Chinese Academy of Sciences and the National Natural Science Foundation of China (30325008 and 30530070).

References

Alexandrov I, Kazakov A, Tumeneva I, Shepelev V, Yurov Y (2001) Alpha-satellite DNA of primates: old and new families. Chromosoma 110:253–266

Amor DJ, Choo KH (2002) Neocentromeres: role in human disease, evolution, and centromere study. Am J Hum Genet 71:695–714

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

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

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 Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691–1704

Clarke L, Carbon J (1985) The structure and function of yeast centromeres. Annu Rev Genet 19:29–55

Cleveland DW, Mao Y, Sullivan KF (2003) Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. Cell 112:407–421

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190

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

Dover GA (2005) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet 2:159–165

Ge S, Sang T, Lu BR, Hong DY (1999) Phylogeny of rice genomes with emphasis on origins of allotetraploid species. Proc Natl Acad Sci USA 96:14400–14405

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Harrison GE, Heslop-Harrison JS (1995) Centromeric repetitive DNA sequences in the genus *Brassica*. Theor Appl Genet 90:157–165

Henikoff S, Dalal Y (2005) Centromeric chromatin: what makes it unique? Curr Opin Genet Dev 15:177–184

Henikoff S, Malik HS (2002) Centromeres: selfish drivers. Nature 417:227

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

Hizume M, Shibata F, Matsusaki Y, Garajova Z (2002) Chromosome identification and comparative karyotypic analyses of four *Pinus* species. Theor Appl Genet 105:491–497

Houben A, Brandes A (1996) Molecular-cytogenetic characterization of a higher plant centromere/kinetochore complex. Theor Appl Genet 93:477–484

Jiang J, Birchler JA, Parrott WA, Kelly DR (2003) A molecular view of plant centromeres. Trends Plant Sci 8:570–575

Kawabe A, Nasuda S (2005) Structure and genomic organization of centromeric repeats in *Arabidopsis* species. Mol Genet Genomics 272:593–602

Kipling D, Warburton PE (1997) Centromeres, CENP-B and Tigger too. Trends Genet 13:141–145

Kishii M, Tsujimoto H (2002) Genus-specific localization of the *Tai*I family of tandem-repetitive sequences in either the centromeric or subtelomeric regions in *Triticeae* species (*Poaceae*) and its evolution in wheat. Genome 45:946–955

Kumar S, Tamura K, Nei M (2004) MEGA 3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Kumekawa N, Ohmido N, Fukui K, Ohtsubo E, Ohtsubo H (2001) A new gypsy-type retrotransposon, RIRE7: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. Mol Genet Genomics 265:480–488

Kuznetsova IS, Prusov AN, Enukashvily NI, Podgornaya OI (2005) New types of mouse centromeric satellite DNAs. Chromosome Res 13:9–25

Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. Proc Natl Acad Sci USA 102:11793–11798

Martinez-Zapater JM, Estelle MA, Somerville CR (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol Gen Genet 204:417–423

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

Murata M, Ogura Y, Motoyoshi F (1994) Centromeric repetitive sequences in *Arabidopsis thaliana*. Jpn J Genet 69:361–370

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet 36:138–145

Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, Cheng Z, Jiang J (2005) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol Biol Evol 22:845–855

Nakajima R, Noma K, Ohtsubo H, Ohtsubo E (1996) Identification and characterization of two tandem repeat sequences (TrsB and TrsC) and a retrotransposon (RIRE1) as genome-general sequences in rice. Genes Genet Syst 71:373–382

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp 1.132–1.142

Sun X, Wahlstrom J, Karpen G (1997) Molecular structure of a functional *Drosophila* centromere. Cell 91:1007–1019

Talbert PB, Bryson TD, Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. J Biol 3:18

Tek AL, Jiang J (2004) The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. Chromosoma 113:77–83

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Ugarkovic D, Plohl M (2002) Variation in satellite DNA profiles—causes and effects. EMBO J 21:5955–5959

Vaughan DA, Morishima H, Kadowaki K (2003) Diversity in the *Oryza* genus. Curr Opin Plant Biol 6:139–146

Willard HF (1991) Evolution of alpha satellite. Curr Opin Genet Dev 1:509–514

Zhang D, Yang Q, Bao W, Zhang Y, Han B, Xue Y, Cheng Z (2005) Molecular cytogenetic characterization of the *Antirrhinum majus* genome. Genetics 169:325–335

Zhao T, Wu T, Xie Y, Wu R (1989) Genome-specific repetitive sequences in the genus *Oryza*. Theor Appl Genet 76:835–840