Funct Integr Genomics (2010) 10:477–492
DOI 10.1007/s10142-010-0190-3

ORIGINAL PAPER

The compact *Brachypodium* genome conserves centromeric regions of a common ancestor with wheat and rice

Lili Qi · Bernd Friebe · Jiajie Wu · Yongqiang Gu · Chen Qian · Bikram S. Gill

Received: 13 July 2010 / Revised: 20 August 2010 / Accepted: 24 August 2010 / Published online: 15 September 2010

Abstract The evolution of five chromosomes of *Brachypodium distachyon* from a 12-chromosome ancestor of all grasses by dysploidy raises an interesting question about the fate of redundant centromeres. Three independent but complementary approaches were pursued to study centromeric region homologies among the chromosomes of *Brachypodium,* wheat, and rice. The genes present in pericentromeres of the basic set of seven chromosomes of wheat and the Triticeae, and the 80 rice centromeric genes spanning the CENH3 binding domain of centromeres 3, 4, 5, 7, and 8 were used as "anchor" markers to identify centromere locations in the *B. distachyon* chromosomes. A total of 53 *B. distachyon* bacterial artificial chromosome (BAC) clones anchored by wheat pericentromeric expressed sequence tags (ESTs) were used as probes for BAC-fluorescence in situ hybridization (FISH) analysis of *B. distachyon* mitotic chromosomes. Integrated sequence alignment and BAC-FISH data were used to determine the approximate positions of active and inactive centromeres in the five *B. distachyon* chromosomes. The following syntenic relationships of the centromeres for *Brachypodium* (Bd), rice (R), and wheat (W) were evident: Bd1-R6, Bd2-R5-W1, Bd3-R10, Bd4-R11-W4, and Bd5-R4. Six rice centromeres syntenic to five wheat centromeres were inactive in *Brachypodium* chromosomes. The conservation of centromere gene synteny among several sets of homologous centromeres of three species indicates that active genes can persist in ancient centromeres with more than 40 million years of shared evolutionary history. Annotation of a BAC contig spanning an inactive centromere in chromosome Bd3 which is syntenic to rice *Cen8* and W7 pericentromeres, along with BAC FISH data from inactive centromeres revealed that the centromere inactivation was accompanied by the loss of centromeric retrotransposons and turnover of centromere-specific satellites during Bd chromosome evolution.

Keywords Centromere · Synteny · *Brachypodium* · Wheat · Rice

Electronic supplementary material The online version of this article (doi:10.1007/s10142-010-0190-3) contains supplementary material, which is available to authorized users.

L. Qi (✉)
Northern Crop Science Laboratory, USDA-ARS,
1605 Albrecht Blvd N Fargo, ND 58102-2765, USA
e-mail: lili.qi@ars.usda.gov

B. Friebe · C. Qian · B. S. Gill
Wheat Genetic and Genomic Resources Center,
Department of Plant Pathology, Kansas State University,
Manhattan, KS 66506-5502, USA

J. Wu · Y. Gu
Genomics and Gene Discovery Research Unit,
Western Regional Research Center, USDA-ARS,
800 Buchanan Street,
Albany, CA 94710, USA

Introduction

Representatives of the grass family Poaceae include economically important crop plants such as maize (1C=2,300 Mb, x=10) and sorghum (1C=730 Mb, x=10) of the subfamily Panicoideae, rice (1C=389 Mb, x=12) of the Ehrhartoideae, and common wheat (1C=17,000 Mb, 3x=21) and barley (1C=5,096 Mb, x=7) of the Pooideae, to which also belongs the new monocot model *Brachypodium distachyon* (1C=272 Mb, x=5) (Bennett and Leitch 2005; International Rice Genome Sequencing Project 2005; Paterson et al. 2009; Schnable et al. 2009; The International Brachypodium Initiative 2010). Fifty to 70 million years of

shared evolutionary history has shaped the diversity of genome sizes and basic chromosome numbers found in these grasses (reviewed by Bolot et al. 2009). The genomes of rice, sorghum, maize, and *Brachypodium* have been sequenced (Goff et al. 2002; Yu et al. 2002; International Rice Genome Sequencing Project 2005; Paterson et al. 2009; Schnable et al. 2009; The International Brachypodium Initiative 2010) and high-density expressed sequence tag (EST) maps are available for wheat (Qi et al. 2004; Luo et al. 2009) and barley (Stein et al. 2007). The rice genome sequence data revealed a whole-genome duplication (WGD) event that happened in a common ancestor of the grasses between 53 and 94 million years ago (Yu et al. 2005). Based on the analysis of shared duplications, $x=5$ was proposed as the ancestral chromosome number of all cereal grasses (Salse et al. 2008). The $x=5$ ancestor, following a WGD, and fission and fusion events involving two chromosomes, produced an intermediate ancestor with $x=12$. The basic chromosome numbers of $x=12$ in rice, $x=10$ in sorghum, $x=7$ in wheat and barley, and $x=5$ in *Brachypodium* trace to this common $x=12$ chromosome ancestor (Salse et al. 2008; The International Brachypodium Initiative 2010). In maize, the 12 chromosomes fused to form a $x=5$ derivative that underwent WGD to produce $x=10$ in maize (Salse et al. 2008).

Centromeres are prominent and pivotal chromosome landmarks, but their fate during this convoluted chromosome evolution has been more difficult to determine, because in most complex eukaryotes studied so far, including *Drosophila melanogaster*, humans, mice, maize (*Zea mays*), rice (*Oryza sativa*), and *Arabidopsis thaliana*, centromeric repeats are among the most rapidly evolving DNA sequences that can differ even between closely related species (Henikoff et al. 2001; Hall et al. 2003; Jiang et al. 2003; Wong and Choo 2004; Henikoff and Dalal 2005; Lee et al. 2005; Sharma and Presting 2008; Birchler et al. 2009; Gao et al. 2009; Wu et al. 2009). This has hindered comprehensive sequence analysis in centromere regions across the organisms. Following decades of research, centromere boundaries, with few exceptions, have been identified for all 12 rice chromosomes and contiguous sequences of the centromeres of rice chromosomes 3, 4, 5, and 8 have been assembled (Cheng et al. 2002; Nagaki et al. 2004; Wu et al. 2004; Zhang et al. 2004a; International Rice Genome Sequencing Project 2005; Yan et al. 2006, 2008). The centromere positions of maize chromosomes were identified by genetic and FISH mapping of the centromere sequence related repeat junction and CENH3 epigenetic markers (Wolfgruber et al. 2009).

The discovery of active genes in functional domains of most rice chromosome centromeres immediately suggested that conserved centromeric gene sequences (COS-C) may be used to study comparative homologies of centromeric regions across the grass species (Nagaki et al. 2004; Wu et al. 2004; Zhang et al. 2004a; Yan et al. 2006, 2008). In a pilot study, we demonstrated that the rice centromeric genes located in the CENH3 functional domain of chromosome 8 were conserved in wheat and mapped to the centromeric regions of group-7 chromosomes of wheat (Qi et al. 2009). Comparative COS-C mapping and bacterial artificial chromosome-fluorescence in situ hybridization (BAC-FISH) landing were used to establish centromere homologies between wheat and rice chromosomes (Qi et al. 2009).

B. distachyon is a wild temperate grass of the Brachypodieae tribe of Pooideae, which contains many important temperate cereals such as wheat, barley, and oat (Shi et al. 1993; Catalan et al. 1995, 1997; Catalan and Olmstead 2000). Close phylogenetic position to temperate cereals, small genome size (~272 Mb), low content of highly repeated DNA (~21.4%), ease of transformation, and short life cycle prompted the use of *B. distachyon* as an experimental model to link rice and temperate grasses (Catalan et al. 1995; Draper et al. 2001; Garvin et al. 2008; Opanowicz et al. 2008; The International Brachypodium Initiative 2010). Comparative analysis using ESTs, BAC-end sequences, and sequenced BAC clones of *Brachypodium* with wheat and rice revealed a closer relationship between *Brachypodium* and wheat than between wheat and rice (Griffiths et al. 2006; Vogel et al. 2006; Bossolini et al. 2007; Huo et al. 2008, 2009; Faris et al. 2008). The genome sequence of *B. distachyon* and its 8× assembly are now publicly available (www.brachypodium.org). Here, we report the COS-C sequence alignment of 36 wheat and 80 rice centromeric genes to the sequenced genome of *Brachypodium* and BAC-FISH landing to identify the putative positions of active and inactive centromeres in the *Brachypodium* genome. Five of the *B. distachyon* putative centromere regions may be traced back to the ancient centromeres of the five ancestral chromosomes and their duplicated homologs of grasses (Salse et al. 2008; Bolot et al. 2009).

Materials and methods

Seeds of *B. distachyon*, an inbred, diploid line Bd21 used for BAC-FISH mapping, were obtained from USDA-ARS, Pacific West Area, Western Regional Research Center, Genomics and Gene Discovery, Albany, CA, USA. Ten BAC clones, previously mapped to five individual *Brachypodium* chromosomes, were provided by Dr. R. Hasterok, University of Silesia, 40-032 Katowice, Poland (Table S1, Hasterok et al. 2006).

Identification of putative centromere BACs of *B. distachyon*

Nucleic acid sequence alignments

The sequences of 35 wheat ESTs previously mapped to the centromere regions and one cDNA clone PSR161 previously mapped to the centromere of chromosome 1B (Sandhu et al. 2001; Francki et al. 2002; Qi et al. 2006, 2009; http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi) were used as queries search against the 4× whole-genome sequences of *B. distachyon* to identify putative centromere BACs (Table 1; http://www.brachybase.org/blast/). These wheat EST sequences were also searched against the *B. distachyon* 8× assembly to align the wheat ESTs to individual *B. distachyon* chromosomes. Centromere synteny between rice and *Brachypodium* was determined using BLASTN searches of the 80 rice active gene model sequences spanning the CENH3 binding domain of rice centromeres 3, 4, 5, 7, and 8 against the *B. distachyon* 8× assembly sequences (Table S3; Yan et al. 2005, 2006, 2008).

BAC library screening

The *B. distachyon Hin*dIII BAC library consisting of 36,864 clones representing 9.9 genome equivalents was developed by Huo et al. (2006). Library screening was performed using one high-density filter containing 18,432 clones (~4.5×coverage); probing with RFLP clone PSR161 identified four *B. distachyon* BAC clones. The procedure for colony filter hybridization was described by Qi et al. (2009).

Computational annotation of a Bd3 BAC contig

A Bd3 BAC contig sequence was first masked for repetitive elements using the RepeatMasker program with the *Oryza* repeat database (http://www.repeatmasker.org/, version 20090604). To annotate long terminal repeat (LTR) retrotransposons in the BAC contig, the *Brachypodium* whole-genome sequence Brachy 1.0 unmasked was downloaded from MIPS (http://mips.helmholtz-muenchen.de/plant/brachypodium/download/index.jsp) and scanned with LTR_STRUC to generate a database of complete LTR retrotransposons (McCarthy and McDonald 2003). BAC sequences were used as queries in BLASTN searches against the database. In addition, Dot Plots alignments were performed to identify possible LTR regions. The BAC sequences were also subjected to tBLASTx searches of the Triticeae Repeat Sequence Database (TREP, http://wheat.pw.usda.gov/ITMI/Repeats/) to identify retrotransposon-coding regions. The types of LTR retrotransposons were determined based on hits in tBLASTx results against TREP. Tandemly repeated elements were identified by using Tandem Repeats Finder program (Benson 1999; http://tandem.bu.edu/trf/trf.html).

BAC-FISH analysis

Mitotic metaphase chromosome spreads were prepared from root tips of *B. distachyon*. Root tips approximately 1.5 to 2 cm long were cut, pretreated with ice water for 24 h, and fixed in ethanol (100%)/glacial acetic acid (3:1). Squash preparations were made in 45% acetic acid, cover slips were removed after freezing on dry ice, and the preparations were then dehydrated in ethanol for 5 min.

The following DNA probes were used in this study: (1) A total of 49 BAC clones of *B. distachyon* selected based on the best hit to wheat centromeric region ESTs (Table 1), (2) four positive BAC clones from the library screening (Table 1), and (3) ten BAC clones with known chromosomal locations as examined by Hasterok et al. (2006) (Table S1). The wheat ribosomal probe pTa71, which contains 18S–5S–26S rDNA sequences (Gerlach and Bedbrook 1979), also was used for in situ hybridization experiments. BAC and plasmid DNA was isolated using a QIAGEN Plasmid Midi Kit 312143 (Qiagen, Valencia, CA.) according to the manufacturer's protocols. One microgram of BAC or plasmid DNA was labeled with fluorescein-12-dUTP (Enzo Life Science Inc, Farmingdale, NY) or Tetramethyl-rhodamine-5-dUTP (Roche Applied Science, Indianapolis, IN) using nick translation. The BAC-FISH procedure was as described by Zhang et al. (2004b). Slides were analyzed with an epifluorescence Zeiss Axioplan 2 microscope. Images were captured using a SPOT 2.1 CCD (charge-coupled device) camera (Diagnostic Instruments) and processed with Photoshop v5.5 (Adobe Systems).

Results

Anchoring wheat pericentromeric ESTs and rice centromeric genes to the *B. distachyon* genome

Three independent but complementary approaches were pursued to study centromeric region homologies among the chromosomes of *Brachypodium* (x=5), wheat (x=7), and rice (x=12). In one approach, genes present in wheat pericentromeres were compared against the sequenced genome of *Brachypodium* (http://blast.brachybase.org/). Thirty-two of 36 wheat cDNA clones in syntenic blocks marking the location of seven wheat and Triticeae (x=7) pericentromeres (Qi et al. 2009) detected homologous syntenic blocks in the four *B. distachyon* chromosomes

Table 1 *Brachypodium* bacterial artificial chromosome (BAC) clones anchored by wheat expressed sequence tags (ESTs) and RFLP clone and their BAC-fluorescence in situ hybridization (FISH) results

Wheat EST or probe	Bin location[a]	*Brachypodium*						
		BAC clone	BES[b]	4× Supercotig	Chromosome	BAC-FISH signal		
						Location	Involved chromosome	Intensity
PSR161	Cen1B	DH063O21[c]	NA	NA	2	Centromere	All	++
		DH096G2[c]	NA	NA	2	Centromere	All	++
		DH053N18[c]	NA	NA	2	Centromere	One pair	++
		DH096P22[c]	NA	NA	2	Centromere	One pair	++
		DH005A16	DH005A16_R1.ab1	4	2	Centromere	All	+
BE500625	C-2AS5-0.78 C-2BL2-0.36	NA	NA	NA	NA	NA	NA	NA
BE404630	C-2BS1-0.53 C-2L-0.49	DB042G22	DB042G22_R1.ab1	1	1	Pericentromere	One pair	++
BF485348	C-3S-0.24 C-3BL2-0.22	NA	NA	2	2	NA	NA	NA
BE637878	C-3BS1-0.33	DB017O12	DB017O12_F1.ab1	0	1	Pericentromere	All	++
	C-3DL2-0.27					Interstitial	One pair	+++
BG313557	C-3BS1-0.33	DB069H16	DB069H16_F1.ab1	0	1	Centromere	All	+
	C-3DL2-0.27	DB041A24	DB041A24_R1.ab1	4	2	Centromere	All	+++
		DH017G05	DH017G05_R1.ab1	4	2	Centromere	All	+++
		DH039C01	DH039C01_R1.ab1	4	2	Centromere	All	+++
BE404580	C-3BS1-0.33	DH010P20	DH010P20_F3.ab1	4	2	Centromere	All	+++
	C-3L-0.27	DH028E07	DH028E07_R1.ab1	4	2	Centromere	All	+
		DH039J16	DH039J16_R1.ab1	4	2	Centromere	All	+++
		DH006E07	DH006E07_R1.ab1	4	2	Centromere	All	+++
		DH039J18	DH039J18_F1.ab1	4	2	Centromere	All	++
		DH035E22	DH035E22_F1.ab1	4	2	–	–	–
		DH046D04	DH046D04_F2.ab1	4	2	Interstitial	One pair	+
BE497309	C-4DS1-0.53 C-4L-0.43	DB159P21	DB159P21_F1.ab1	1	1	Centromere	All	+
BE497635	C-4DS1-0.53	DH034K22	DH034K22_F1.ab1	2	2	Centromere	All	+++
	C-4L-0.43					Distal	One pair	++
BE406512	C-4S-0.37 4DL9-0.31-0.56	NA	NA	NA	1	NA	NA	NA
BF202969	C-4BS4-0.37 C-4L-0.31	DH004K12	DH004K12_F1.ab1	1	1	Centromere[d]	All	++

BE494281	C-4BS4-0.37 C-4L-0.31	DH018M04	DH018M04_F1.ab1	10	5	Centromere[d]	All	++
BF202706	C-4BS4-0.37 C-4L-0.31	DH018E21	DH018E21_F1.ab1	6	4	Centromere	All	+
BE637507	C-4L-0.31	DB069J23	DB069J23_R1.ab1	15	4	Centromere	All	+++
		DH024F10	DH024F10_R8.ab1	15	4	Centromere	All	++
		DB029E18	DB029E18_F1.ab1	15	4	Centromere	All	++
		DB048G17	DB048G17_F1.ab1	15	4	Centromere	All	+++
		DB081A17	DB081A17_F1.ab1	15	4	–	–	-
		DB020G08	DB020G08_R1.ab1	15	4	Centromere	All	+++
BE403618	C-5S-0.40	DB071N11	DB071N11_R1.ab1	7	4	Distal	One pair	+
	C-5AL12-0.35	DB091F03	DB091F03_F1.ab1	7	4	–	–	-
		DH008K14	DH008K14_R1.ab1	7	4	Centromere	All	++
		DH047L23	DH047L23_F1.ab1	7	4	Distal	One pair	++
		DH044G07	DH044G07_F1.ab1	7	4	Distal	One pair	+
BG263528	C-5AS1-0.40 C-5L-0.60	DH016H07	DH016H07_F1.ab1	7	4	–	–	–
BM140334	C-5AS1-0.40 C-5L-0.29	DB060C06	DB060C06_R1.ab1	7	4	Distal	One pair	++
BF291333	C-5AS1-0.40 C-5L-0.29	DB017D18	DB017D18_F1.ab1	7	4	–	–	–
BE497510	C-5AS1-0.40 C-5L-0.29	DB070M14	DB070M14_R1.ab1	7	4	Distal	One pair	++
BG263803	C-5AS1-0.40 C-5L-0.29	DH037A19	DH037A19_R1.ab1	7	4	Distal	One pair	++
BE406602	C-6S-0.35	DH008E08	DH008E08_R1.ab1	3	3	NA	NA	NA
	C-6BL3-0.36	DH022F20	DH022F20_R1.ab1	3	3	NA	NA	NA
BE405809	C-6DS2-0.45 C-6BL3-0.36	DB042E22	DB042E22_F1.ab1	3	3	Centromere	All	+++
BE405195	C-6S-0.35	DB088O14	DB088O14_F1.ab1	5	3	Centromere	All	+++
	C-6BL3-0.36	DB145B13	DB145B13_F1.ab1	5	3	Centromere	All	+++
		DB164M07	DB164M07_F1.ab1	5	3	Centromere	All	+
		DB089K04	DB089K04_R1.ab1	5	3	Centromere	All	++
		DB088K03	DB088K03_R1.ab1	5	3	Centromere	All	+
BF428553	C-6BS5-0.76	DB025G19	DB025G19_F1.ab1	5	3	Centromere	All	++
	C-6L-0.29	DB083H01	DB083H01_R1.ab1	5	3	Centromere	All	+

Table 1 (continued)

Wheat EST or probe	Bin location[a]	*Brachypodium*				BAC-FISH signal		
		BAC clone	BES[b]	4× Supercotig	Chromosome	Location	Involved chromosome	Intensity
		DH033I11	DH033I11_F1.ab1	5	3	Centromere	All	+
BJ301191	C-7S-0.15	DH010E19	DH010E19_R1.ab1	3	3	Pericentromere	One pair	++
BJ305475	C-7S-0.15	DH010E19	DH010E19_R1.ab1	3	3	Pericentromere	One pair	++
		DB044I06	DB044I06_F1.ab1	3	3	Centromere	All	+++
BJ280500	C-7DS-0.15	DB161L06	DB161L06_F1.ab1	3	3	Pericentromere	One pair	++
	C-7L-0.14	DH030D19	DH030D19_R1.ab1	3	3	Centromere	All	+++
		DB024G07	DB024G07_R1.ab1	3	3	Centromere	All	+++

–, and +represent the absence and presence of hybridization signals, respectively: +++, strong signal; ++, intermediate signal; +, weak signal; NA, not available

[a] Mapping data were taken from http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi and Qi et al. (2006, 2009)

[b] BAC end sequence (BES)

[c] *Brachypodium* BAC clone was obtained from BAC library screening

[d] FISH single also painted *B. distachyon* chromosomes over their entire length

(Fig. 1, Table S2). In the second approach, we explored the centromere synteny between *Brachypodium* and rice by searching the 80 rice active genes spanning the CENH3 binding domain of *Cen3*, *Cen4*, *Cen5*, *Cen7*, and *Cen8* against the JGI 8× *B. distachyon* genome assembly (Yan et al. 2005, 2006, 2008; http://blast.brachybase.org/). Of the 80 rice centromeric active genes, 19 lie on *Cen3*, 17 on *Cen4*, 14 on *Cen5*, 18 on *Cen7*, and 12 on *Cen8* (Table S3). The best matches of *B. distachyon* sequences to active rice genes in *Cen3*, *Cen4*, *Cen5*, *Cen7*, and *Cen8* were 47%, 59%, 57%, 56%, and 67%, respectively (Table 2). The rice centromere syntenic blocks detected homologous syntenic blocks on all five *Brachypodium* chromosomes (Fig. 1). In the final approach, BAC-FISH landing was used to correlate the location of homologous syntenic blocks to physical centromeres of *B. distachyon* chromosomes (Table 1, Figs. 1 and 2). These results are summarized below for the five *Brachypodium* chromosomes.

Chromosome 1 (Bd1)

Bd1 is the largest chromosome at 75 Mb, is submetacentric, and can be distinguished from the other chromosomes based on size alone (Draper et al. 2001). Four out of six (4/6) centromeric genes from the CENH3 domain of rice centromere 3 (R3-C) gave a hit at 11.0–13.0 Mb of Bd1. Evidently, this is the location of an inactive centromere (Fig. 1). A subset of genes (3/4) from the wheat group-4 (W4) pericentromeric region aligning to R3 gave a hit at 59.8–62.0 Mb of Bd1. BAC-FISH of Bd-BACs corresponding to W4 ESTs BE497309 and BF202969, with hits at 59.8–60.7 Mb on Bd1, gave a centromeric signal on all Bd centromeres, indicating that repetitive sequences at the Bd1 59.8–60.7 Mb site are enriched in centromeres of all *Brachypodium* chromosomes (Table 1).

Another syntenic block of four centromeric genes from R7 CENH3 domain gave a hit at position 49.8–51.0 Mb on Bd1 (Fig. 1). Four of the five pericentromere genes of W2 that aligned adjacent to R7-C in the short arm also aligned adjacent to the R7-C-homologous block in Bd1 at 51.9–53.5 Mb. BAC-FISH of a Bd-BAC corresponding to W2-BE404630 gave a hit only on Bd1 at 53.5 Mb; the FISH signal was proximal to the Bd1 centromere (Fig. 2c). In addition, we searched a rice *Cen6* BAC clone OSJN-Ba0015G09 encompassing 11 rice pseudomolecules against the *Brachypodium* genome sequence. Out of 11 rice genomic DNA sequences, three putatively expressed genes gave a hit at 38.4–38.7 Mb of Bd1 (Table S4, Fig. 1). This is possibly an active centromere in Bd1 based on the morphology of Bd1 and annotation of *Brachypodium* centromeric repeats (BdCENT) in this region (The International Brachypodium Initiative 2010).

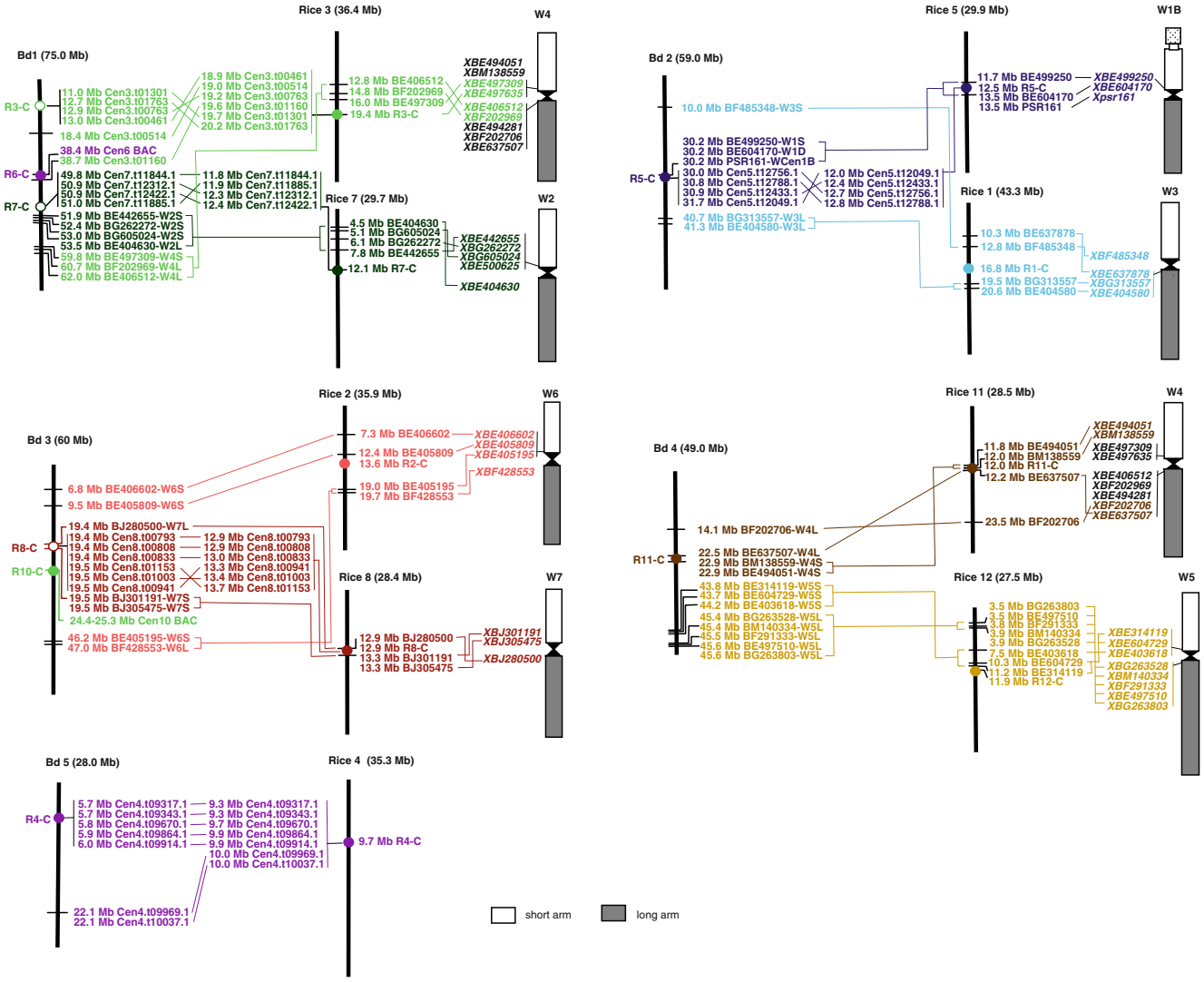

Fig. 1 Comparative mapping of wheat pericentromeric ESTs in rice and *B. distachyon* and the rice centromeric genes of *Cen3*, *Cen4*, *Cen5*, *Cen7*, and *Cen8* in *B. distachyon* by sequence BLASTN search. The centromere positions indicated by *solid circles* in rice chromosomes were taken from http://rice.plantbiology.msu.edu/pseudomolecules/centromere.shtml. A *solid circle* in a *B. distachyon* chromosome represents an active rice centromere, and an *open circle* in a *B. distachyon* chromosome represents an inactive rice centromere. Probes with homologous sequence are represented with the *same color*

Table 2 Summary of active rice genes spanning the CENH3 binding domain of *Cen3*, *Cen4*, *Cen5*, *Cen7*, and *Cen8* aligning to *B. distachyon* chromosomes

Rice	No. of active genes	Rice genes aligning to Bd chromosome					
		Total	Bd1	Bd2	Bd3	Bd4	Bd5
Cen3	19	9 (47)	7	1		1	
Cen4	17	10 (59)	1	1		1	7
Cen5	14	8 (57)	2	5	1		
Cen7	18	10 (56)	6	1	1		2
Cen8	12	8 (67)	1		6	1	

The number in parentheses represents the percentage of rice centromeric genes syntenic to *Brachypodium* chromosomes

Chromosome 2 (Bd2)

Chromosomes Bd2 and Bd3 are of similar size, each ca. 59–60 Mb in size, but Bd2 is metacentric and Bd3 is submetacentric. Genes mapping to the W1, R5, and Bd2 centromeres had perfect homology in centromere position and conservation (Fig. 1). The three genes mapping to the W1 pericentromere mapped to R5-C; these three W1 genes and another four rice genes located in the CENH3 domain of R5-C detected a syntenic block at 30.2–31.7 Mb in Bd2. Two of the five Bd-BAC clones detected by PSR161 that was located in the centromere of wheat chromosome 1B gave a BAC-FISH signal only at the Bd2 centromere, and the other three had BAC-FISH signals at the centromeres of

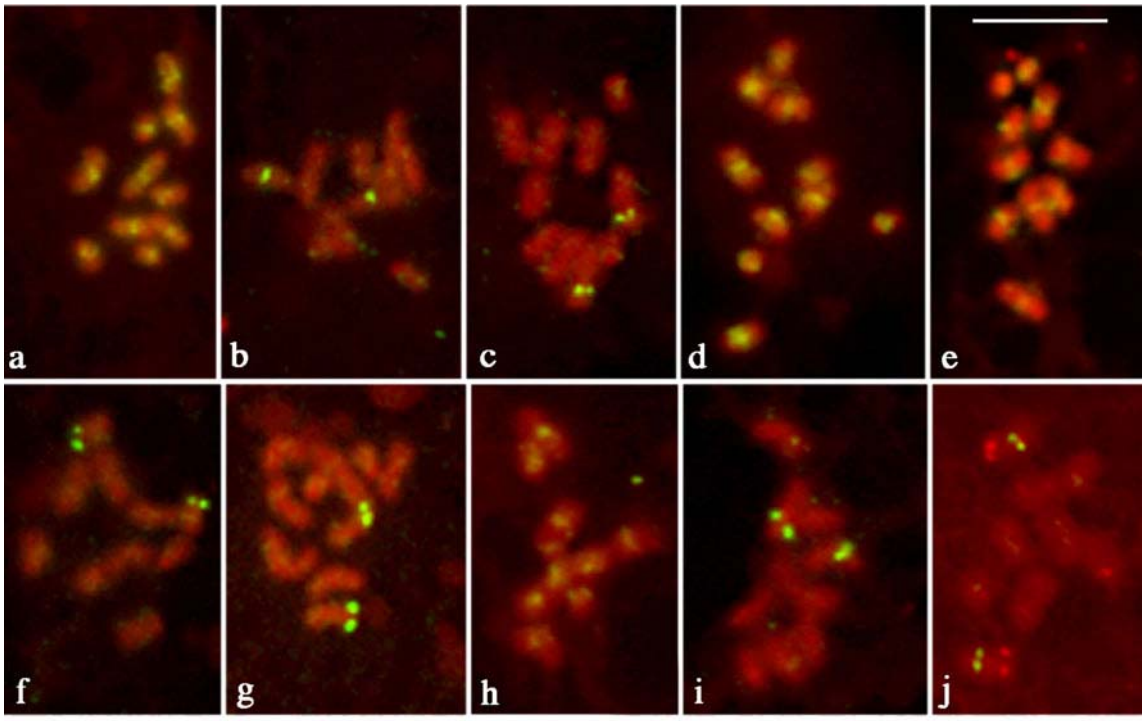

Fig. 2 Mitotic metaphase fluorescence in situ hybridization (FISH) patterns of *B. distachyon* chromosomes using *B. distachyon* BAC clones anchored by wheat ESTs as probe. The signal was visualized by yellow-green FITC florescence and *B. distachyon* chromosomes were counterstained with propidium iodide and fluoresce red. **a** BAC DH063O21 corresponding to wheat Cen1B-PSR161; the signal is in all centromeres. **b** BAC DH053N18 corresponding to wheat Cen1B-PSR161; the signal is in the centromeres of a pair of chromosomes. **c** BAC DB042G22 corresponding to W2L-BE406430; the signal is in the pericentromeric region of one chromosome pair. **d** BAC DH039C01 corresponding to W3L-BG313557, the signal is in all centromeres. **e** BAC DH024F10 corresponding to W4L-BE637507; the signal is in all centromeres. **f** BAC DB070M14 corresponding to W5L-BE497510; the signal is in the distal region of one chromosome pair. **g** BAC DH037A19 corresponding to W5L-BG263803; the signal is in the distal region of one chromosome pair. **h** BAC DB088O14 corresponding to W6S-BE405195; the signal is in all centromeres. **i** BAC DB161L06 corresponding to W7L-BJ280500; the signal is in the pericentromeric region of one chromosome pair. **j** Multi-FISH, the BAC clone DH010E19 corresponding to W7S-BJ301191 gave a FISH signal in the pericentromeric region of one chromosome pair (visualized by yellow-green FITC fluorescence), the reference BAC clone ABR1-56H06 that previously mapped to the distal region of the long arm of the chromosome Bd3 co-localized with DH010E19 (visualized by red rhodamine fluorescence). Bar=10 μm

all Bd chromosomes (Fig. 2a and b, Table 1). In the other syntenic block of W3-R1-Bd2, four genes mapping to the W3 pericentromere and spanning the short and long arms of R1-C, mapped to the short and long arms of chromosome Bd2. Bd-BACs detected by wheat W3 pericentromeric clones BG313557 and BE404580, which form a syntenic block at 41 Mb of Bd2, gave BAC-FISH signals at the centromeres of all Bd chromosomes (Fig. 2d). This region of Bd2 may represent the location of an inactive centromere. One of the Bd-BACs, DH046D04, gave an interstitial signal only on Bd2 and may represent the physical location of the BE404580 in Bd2 (Table 1).

Chromosome 3 (Bd3)

This chromosome was postulated to have evolved from three rice chromosomes, R2, R8, and R10 (The International Brachypodium Initiative 2010). The centromere synteny between R2 and W6 and R8 and W7 were established previously (Qi et al. 2009). Two wheat ESTs (BE405195 and BF428553) spanning the W6 pericentromeric region, which detected a syntenic block at 19 Mb in the rice R2 long arm, also detected a syntenic block at position 47 Mb in the long arm of Bd3. The other two ESTs mapping in the W6 short arm pericentromeric region also mapped to the short arms of R2 and Bd3 (Fig. 1). All the Bd BAC clones corresponding to three of the four genes that mapped to the W6 pericentromere gave BAC-FISH signals at the centromeres of all Bd chromosomes, indicating the presence of centromere-rich sequence in the Bd3 site where the W6 pericentromeric ESTs resided, (Fig. 2h, Table 1).

In the W7-R8-Bd3 syntenic block, three clones mapping to the W7 pericentromere correspond to the R8 centromere, and these genes, along with six rice CENH3 domain genes located in *Cen8*, gave a hit at the 19.4–19.5 Mb position in Bd3. Two of five Bd-BAC clones anchored by W7 pericentromeric ESTs, DB161L06 and DH010E19, gave a

FISH signal proximal to the Bd3 centromere only (Fig. 2i, Table 1). The other three clones, DB044I06, DH030D19 and DB024G07, hybridized to all Bd centromeres (Table 1). In a multi-FISH experiment, using BAC clone DH010E19 as a probe combined with different reference BACs with known chromosomal locations (Table S1; Hasterok et al. 2006), this BAC colocalized with BAC clone ABR1-56H6 that was previously mapped to the long arm of Bd3. The DH010E19 signal was close to the centromere and an ABR1-56H6 signal was in the distal end (Fig. 2j). Rice *Cen8* may be an inactive centromere in *Brachypodium* chromosome 3.

Rice R10 centromere lacks genes for comparative COS-C mapping. We searched a rice *Cen10* BAC OSJNBa0034E23 sequence against the sequenced genome of *Brachypodium* (http://blast.brachybase.org/). Of the eight genomic clones in this BAC, six (75%) aligned to *Brachypodium* chromosome 3; three were positioned at 10.6 Mb, and the positions of the other three ranged from 24.4 to 29.8 Mb (Table S4). The latter matched the predicted Bd3 centromere position based on the BdCENT distribution in the Bd3 chromosome, indicating that the R10 centromere is probably active in chromosome Bd3 (Fig. 1) (The International Brachypodium Initiative 2010).

Chromosome 4 (Bd4)

Bd4 is a 49-Mb metacentric chromosome (Draper et al. 2001). Three of the four ESTs spanning the W4 pericentromere mapped to R11-C and gave a hit at the expected midposition at 22.5 Mb in Bd4. BAC-FISH of Bd-BACs detected by cDNA clone BE637507 mapping to the W4, R11, and Bd4 centromeres gave strong hybridization signals at the centromeres of all Bd chromosomes (Fig. 2e). The active centromere of Bd4 appears to be homologous to the R11 and W4 centromeres. W5 pericentromere probes, with hits on the R12 short arm, and R12-C gave hits at 44–46 Mb of the Bd4 long arm, which appears to be the location of an inactive centromere homologous to R12 and W5 centromeres. BAC-FISH of all probes with hits to pericentromeres of W5, R12-C, or the short arm of R12 and the Bd4 long arm gave a BAC-FISH signal on one pair of chromosomes at a distal location except for one, Bd BAC DH008K14 (Fig. 2f and g). This Bd BAC clone detected by BE403618 gave a signal at all Bd centromeres (Table 1). A multi-FISH experiment colocalized DH037A19 with the reference BAC clone ABR5-32C1 in the distal end of the Bd4 long arm (data not shown).

Chromosome 5 (Bd5)

The smallest *Brachypodium* chromosome is acrocentric with a 45 S locus (nucleolus organizing region) in the short arm (Draper et al. 2001). Consistent with the acrocentric morphology, five out of seven genes located in the CENH3 domain of R4-C gave hits at 6 Mb of Bd5, which is obviously the position of an active centromere. As reported previously, none of the clones mapping to wheat pericentromeres detected homology to the R4 centromere (Qi et al. 2009).

Annotation of a Bd3 BAC contig sequence syntenic to rice *Cen8*

Three of five BACs in the Bd3 BAC contig landed on active (visible primary constriction) centromeres on all Bd chromosomes as revealed by BAC-FISH, providing an opportunity for identifying structural DNA elements common to all Bd centromeres. The full BAC sequence was retrieved from anchoring a BAC end sequence to the *B. distachyon* genome sequence. The sequence map of this contig spans 286,664 bp starting at 19,323,633 bp in Bd3 pseudomolecules with an order of BAC clones DH030D19 (121 kb), DB024G07 (113 kb), DB044I06 (143 kb), DB161L06 (99 kb), and DH010E19 (116 kb) (Fig. 3b). This BAC contig contains two gaps totaling 2,070 bp (0.7% of the contig sequence), one at position 19,407,326 to 19,407,426 bp (100 bp) and one at position 19,411,720 to 19,413,690 bp (1,970 bp) (Fig. 3b). Sixteen genes were identified in the contig (Fig. 3b; http://www.brachybase.org/cgi-bin/gbrowse/brachy8/). These genes were concentrated in a region of 19,406,227 to 19,584,368 bp (total 178,141 bp length) with a gene density of 1 gene per 11 kb (Table 3, Fig. 3b). Sixteen Bd3 genes matched 14 rice genes that were distributed on six different rice chromosomes (Table 3). Out of 14 rice genes, nine (64%) were located on rice R8 including six rice *Cen8* active genes (Table 3, Fig. 3b). These *Cen8* active genes embedded within the 750-kb CENH3 binding domain (Nagaki et al. 2004; Yan et al. 2005) were mapped to the 85-kb region of Bd3 with a conserved gene order similar to that in rice R8 except the position of gene *Cen8t00941* was switched with that of gene *Cen8t01153* in Bd3 (Fig. 3b and c).

The abundance and distribution of repetitive DNA across the Bd3 BAC contig were first revealed using the RepeatMasker program (http://www.repeatmasker.org/). The 23,366 bp of DNA, or 8.7% of the 286,664 bp contig sequence, consists of repetitive DNA, the majority belonging to retrotransposons (84.2%), followed by DNA transposons (9.9%), and unknown repeats (5.9%) (Table 4). Up to 66.2% of the retrotransposons belong to the LTR-*gypsy* group, 93% of LTR-*gypsy* DNA was distributed in the first 80,000 bp of the Bd3 contig sequence where no gene was present (Fig. 3). The first three Bd3 BAC clones, DH030D19, DB024G07, and DB044I06, contain 93.3%, 83.3%, and 56.7% of LTR-*gypsy* sequences, respectively.

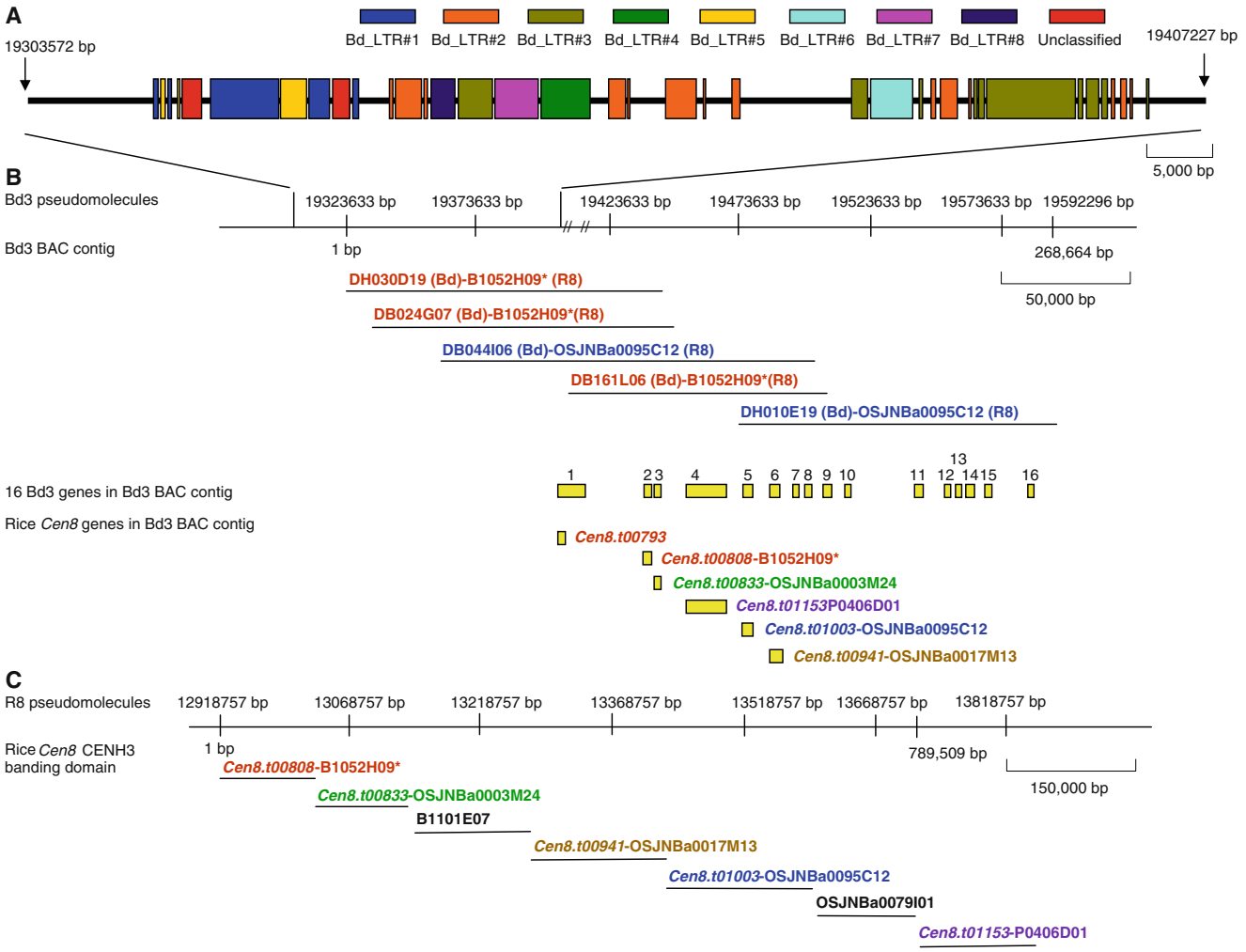

Fig. 3 Map of a Bd3 BAC contig syntenic to rice *Cen8*. **a** LTR retrotransposon annotation of 103,655-bp sequence (1–82,594 bp of the BAC contig and 20,061 bp upstream of the contig). The LTR retrotransposons in this region matched eight different LTR elements (color-coded) identified with LTR_STRUC in *Brachypodium* genome. **b** Location of Bd3 BAC clones and genes, as well as rice *Cen8* active genes in the BAC contig. Each Bd BAC clone links with corresponding rice BAC clone. Probes of the *same color* facilitate the identification of homologous sequences between Bd3 BAC contig and rice *Cen8*. The *yellow bars* represent the gene (http://www.brachybase.org/cgi-bin/gbrowse/brachy8/). Two sequence gaps in the contig are marked by a *double slash* (//). **c** The rice *Cen8* CENH3 binding domain (Yan et al. 2005; http://rice.plantbiology.msu.edu/pseudomolecules/ordered_bac_8.shtml). *rice *Cen8* BAC

The sequence data are consistent with our BAC-FISH results. The Ty3-*gypsy* retrotransposon is highly specific to the centromeric regions of grass chromosomes (Presting et al. 1998; Kumar and Bennetzen 1999; Jiang et al. 2003; Nagaki et al. 2005). The three LTR-*gypsy* rich BAC clones exclusively hybridized to the centromeres of all *B. distachyon* chromosomes, whereas two BAC clones, DB161L06 and DH010E19 with none or a few LTR-gypsy elements, gave FISH signals close to the centromere of the Bd3 chromosome only (Table 2, Figs. 2i and j, and 3b).

We further annotated retrotransposons in the 103,655-bp sequence including 1–82,594 bp-rich LTR-*gypsy* elements of the Bd3 BAC contig and 20,061 bp upstream of the contig (Fig. 3a). A total of 56,422 bp, representing 54.4% of the total sequence, were identified as LTR retrotransposon regions. Of those retrotransposons, 83.1% were *gypsy* type, 10.5% were *copia* type, and the rest remained undetermined. The retrotransposon elements in the Bd3 BAC contig match to eight different *Brachypodium* LTR retrotransposons identified with LTR_STRUC in the genome (Bd_LTR#1 to Bd_LTR#8, Fig. 3a). Among the eight LTR retrotransposons, seven belonged to *gypsy* and one to *copia* (Bd_LTR#6). When the LTR sequences of these LTR retrotransposons were searched against the complete genome sequence, the retrotransposons appear to have considerable variation in density along the chromosomes. Within the annotated region, the most abundant element (Bd_LTR#3) accounted for 31.6% of the retrotransposon sequences identified. This *gypsy*-type element was found in two ~7-kb fragments compared to the 11-kb full-length

Table 3 BLASTN results of the 16 genes in the Bd3 19,323,363–19,592,296 contig against the rice genome sequence

Brachypodium genes				Rice genes								
Gene no.	Gene name	Position (bp)	Length (bp)	Chr.	G-DNA	Position (bp)	Length (bp)	*Cen8* genes	BAC clones	Hit score	E-value	Protein
1	*Bradi3g20420.1*	19406227–19418511	12,285	8	Os08g21700	12998789–13003299	4,815	*Cen8.t00793*	B1052H09	961	2.10E-74	DCL, chloroplast precursor, putative, expressed
2	*Bradi3g20430.1*	19437642–19441692	4,051	8	Os08g21760	13026361–13031955	6,161	*Cen8.t00808*	B1052H09	1,568	8.60E-107	Rer1 protein, putative, expressed
3	*Bradi3g20440.1*	19443709–19447389	3,681	8	Os08g21840	13073371–13070340	3,425	*Cen8.t00833*	OSJNBa0003M24	1,325	6.30E-123	50 S ribosomal protein L15, putative, expressed
4	*Bradi3g20450.1*	19455251–19473329	18,079	8	Os08g22864	13738055–13762723	25,051	*Cen8.t01153*	P0406D01	2,062	0	sec8 exocyst complex component specific domain containing protein, expressed
5	*Bradi3g20460.1*	19475609–19480373	4,765	8	Os08g22354	13469493–13476495	7,461	*Cen8.t01003*	OSJNBa0095C12	3,115	0	Polyadenylate-binding protein, putative, expressed
6	*Bradi3g20470.1*	19485044–19490852	5,809	8	Os08g22149	13356676–13364709	8,633	*Cen8.t00941*	OSJNBa0017M13	1,439	4.40E-119	CBS domain containing membrane protein, putative, expressed
7	*Bradi3g20480.1*	19494538–19497087	2,550	12	Os12g15460	8822172–8824514	3,814		OSJNBa0036A15	887	4.40E-40	Pentatricopeptide, putative, expressed
8	*Bradi3g20490.1*	19497927–19502062	4,136	8	Os08g22200	13390384–13396125	5,945		OSJNBa0017M13	791	1.50E-71	Expressed protein
9	*Bradi3g20500.1*	19504902–19510825	5,924	9	Os09g03600	1779309–1784784	6,046		P0415D04	1,379	2.60E-75	Immunoglobulin/major histocompatibility complex, putative, expressed
10	*Bradi3g20510.1*	19512806–19514849	2,044	9	Os09g03600	1779309–1784784	6,046		P0415D04	2,892	1.90E-206	Immunoglobulin/major histocompatibility complex, putative, expressed
11	*Bradi3g20520.1*	19541097–19546704	5,608	5	Os05g49970	28592293–28595447	6,634		OJ1268_B08	3,605	0	Translation initiation factor IF-2, chloroplast precursor, putative, expressed
12	*Bradi3g20530.1*	19550452–19552932	2,481	5	Os05g49970	28592293–28595447	6,634		OJ1268_B08	3,605	0	Translation initiation factor IF-2, chloroplast precursor, putative, expressed
13	*Bradi3g20540.1*	19555617–19557187	1,571	8	Os08g23790	14401863–14403545	2,543		OJ1136_A10	1,941	4.00E-148	Polygalacturonase, putative, expressed
14	*Bradi3g20550.1*	19558456–19564513	6,058	8	Os08g23780	14395267–14399975	5,151		OJ1136_A10	5,035	0	Glycosyl transferase 8 domain containing protein, putative, expressed
15	*Bradi3g20560.1*	19568104–19572511	4,408	6	Os06g07190	3439374–3441243	2,973		P0680A03	669	7.20E-97	RNA polymerase Rpb7, N-terminal domain containing protein, expressed
16	*Bradi3g20570.1*	19582568–19584368	1,801	2	Os02g29810	17729224–17730320	1,097		P0654A08	681	1.00E-24	Conserved hypothetical protein

Table 4 Distribution of repeats along the 268,663-bpBd3 BAC contig

Feature	No. of elements[a]	Length occupied (bp)	Percentage of sequence (%)
Retroelements	41	19,668	7.32
SINEs	1	215	0.08
LINEs	2	1,964	0.73
L1/CIN4	2	1,964	0.73
LTR elements	38	17,489	6.51
Ty1/Copia	6	4,051	1.51
Gypsy/DIRS1	26	13,012	4.84
DNA transposons	12	2,312	0.86
Tc1-IS630-Pogo	2	258	0.10
En-Spm	4	961	0.36
MuDR-IS905	3	677	0.25
Tourist/Harbinger	2	306	0.11
Total interspersed repeats		21,980	8.18
Simple repeats	9	359	0.13
Low complexity	30	1,027	0.38

[a] Most repeats fragmented by insertions or deletions have been counted as one element

elements, one solo LTR, and several deletion derivatives (Fig. 3a). The second most abundant element (Bd_LTR#2) accounted for 23.6% of the retrotransposon sequences. Five copies of this partial *gypsy* type LTR element were in this region. Four of the five copies showed the same structures, resulting from recombination in the LTR regions between two elements followed by deletions and duplications. Alignment of the rearranged partial retrotransposon resulted in matches only to the 5′ and 3′ regions of a full-length LTR retrotransposon (Fig. S1).

The Bd3 BAC contig sequence was aligned to the conserved centromeric retrotransposons (CRs) identified in *B. sylvaticum* (CCS1), wheat (CRW), rye (bilby), barley (ceraba), rice (CRR), sorghum (pHind22), and maize (CRM) (Aragón-Alcaide et al. 1996; Jiang et al. 1996; Ananiev et al. 1998; Dong et al 1998; Miller et al. 1998; Francki 2001; Cheng et al. 2002; Hudakova et al. 2001; Nagaki et al. 2004; Liu et al. 2008). No significant sequence similarities were found between the Bd3 BAC contig with the CR elements of these species, indicating divergent centromere-rich retrotransposons in *Brachypodium*.

Thirty-six tandem repeats were identified in the Bd3 BAC contig using the Tandem Repeat Finder program (Table S5, http://tandem.bu.edu/trf/trf.html). In contrast to the Ty3-*gypsy* retrotransposons concentrated in the first 80,000 bp in the Bd3 BAC contig, most tandem repeats (about 81%, 29/36) were distributed from 81,603 bp to the end of the contig (Table S5). Out of 36 tandem repeats, 34 had a repeat size smaller than 100 bp with a range of copy numbers from 1.9 to 28.5. Two tandem repeats had 128 and 211 bp with a copy number of 2 and 3, respectively. *Brachypodium* centromere-specific satellite repeat of BdCENT was not found in the contig (The International Brachypodium Sequencing Initiative 2010).

Discussion

During chromosome evolution leading to changes in basic chromosome number, centromeres may be conserved, be inactivated, fuse or disperse following a split, change in position, or be eliminated. In the *Arabidopsis* relatives, unequal translocations leading to the elimination of the smaller translocation chromosome have contributed to changes in basic chromosome number from $x=8$ to $x=5$ and in centromere reduction (Kawabe et al. 2006; Lysak et al. 2006). Han et al. (2009) reported that in cucurbit species, a gain/loss of a large amount of pericentromeric heterochromatin were associated with centromere activation and inactivation during evolution of cucurbit chromosomes. In cereals, comparative bioinformatics analysis of sequenced genomes and mapped sequences have indicated that dysploidy, where one whole chromosome is incorporated into the centromere of another chromosome, has played a dominant role in chromosome number reduction (Luo et al. 2009; The International Brachypodium Initiative 2010). In *Brachypodium*, chromosomes Bd1, Bd3, and Bd4 were involved in two nested, dysploid events, Bd2 in a dysploid event, and only Bd5 chromosome has maintained complete colinearity with one rice chromosome (The International Brachypodium Initiative 2010); the positions of the inactive centromeres were inferred from the distribution of repeat families. Our present study provided experimental evidence for this hypothesis and a framework of genes and sequences that mark the positions of active and inactive centromeres.

Chromosome Bd1 harbors centromere homologies to rice chromosomes R3, R6, and R7 and to wheat chromosomes W4 and W2. Centromere sequences related to R3 and W4 were mapped to the distal positions in the short and long arms of Bd1. R7-W2 related sequences were closely

localized in two positions in the long arm of Bd1. This data supports the first insertional dysploidy event where an ancestral R7 was inserted into the R3 centromere (The International Brachypodium Initiative 2010). The active centromere of Bd1 most likely was derived from R6 that was inserted close to the R7 centromere. (The International Brachypodium Initiative 2010).

In chromosome Bd2, the active centromere is derived from R5, which is syntenic to W1 centromeres that mapped at the same positions as the rice centromere R5 in Bd2. Furthermore, our data support the insertion of R5 into the R1 centromere (The International Brachypodium Initiative 2010) because R1-C-W3 specific probes at 40 Mb of Bd2 gave BAC-FISH signals on all Bd centromeres (Fig. 2d).

Two, nested, dysploid events at the centromeres were postulated during the origin of chromosome Bd3 (The International Brachypodium Initiative 2010). Probes marking the R2 and W6 pericentromeric regions were mapped to the distal ends of the long and short arms of Bd3, whereas the R8/W7 centromeres mapped at 19.4 Mb. Thus, R8 was inserted into the centromere of R2. The postulated insertion of R10 into R8 may have occurred because the R10 centromere appears to be active in chromosome Bd3.

Chromosome Bd4 was postulated to have evolved from two, nested, dysploid events, the first involving the insertion of R9 into the centromere of R12 and the second involving the insertion of R11 into the centromere of R9. Our results agree that the Bd4 active centromere shares homology with R11; we did not see any evidence for dysploidy events involving the R11 centromere.

Chromosomes Bd5 and R4 are colinear and share centromere homology as well, but we do not have any data on the conservation of this centromere in the wheat lineage.

In terms of wheat and rice centromere homologies, our results agree with those of Luo et al. (2009), except that our data indicate that the W1 centromere is homologous to R5 and not R10. Three W1 cDNA clones, PSR161 previously mapped to the 1B centromere (Sandhu et al. 2001; Francki et al. 2002), BE604170 to 1D (Luo et al. 2009), and BE499250 to the pericentromeric regions of group-1 chromosomes (http://wheat.pw.usda.gov/cgi-bin/westsql/map_locus.cgi), all mapped to the R5 centromeric region, 0.8–1 Mb flanking R5 centromeric BAC (Fig. 1, Table S2; see also Qi et al. 2009). Thus, during the dysploid reduction to $x=7$ in wheat and the Triticeae from an ancestor with 12 chromosomes, four ancestral centromeres with homologies to R4, R6, R9, and R10 were lost. Furthermore, W4 may be a hybrid centromere with homologies to the R3 and R11 centromeres (Figs. 1 and 4; Qi et al. 2009).

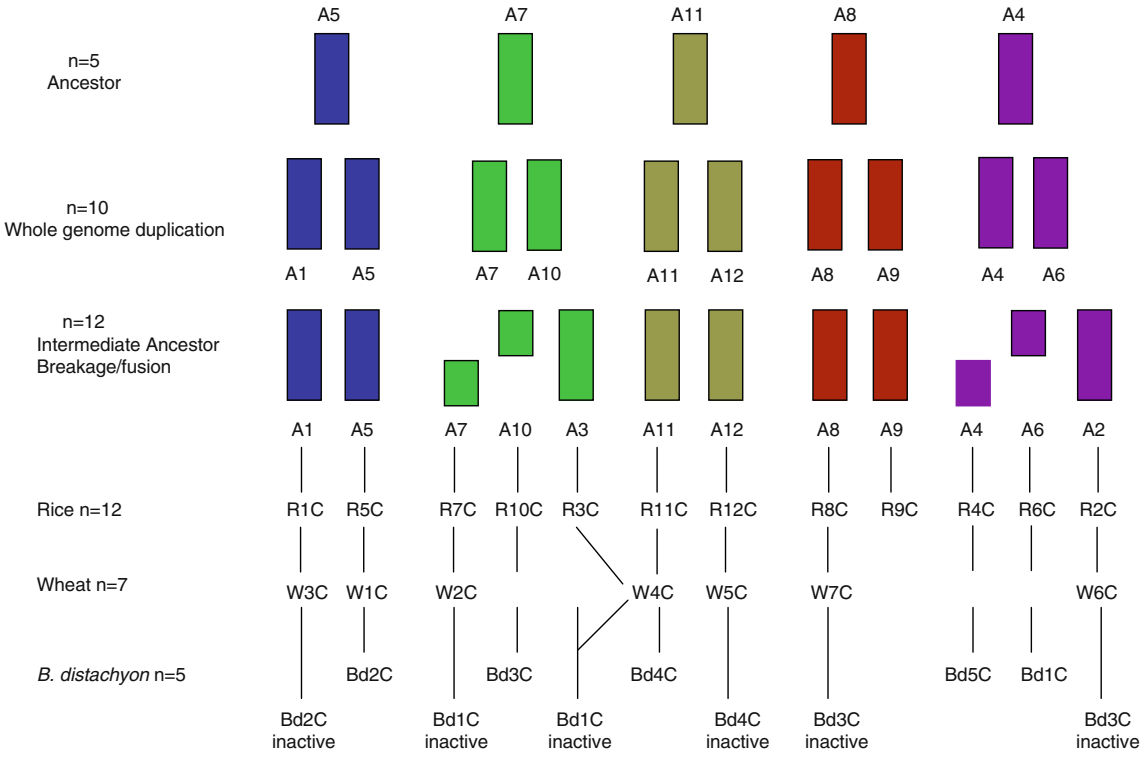

Fig. 4 A model for centromere evolution among rice, wheat, and *Brachypodium* from a common ancestor with $n=5$ chromosomes, modified from Salse et al. (2008). The active centromeres of all *B. distachyon* chromosomes can be traced to the centromeres of either five ancestral chromosomes or their duplicated homologs. Six rice centromeres, syntenic to five wheat centromeres, were inactive in four *Brachypodium* chromosomes. C, centromere

We hypothesize that conserved C-gene syntenic blocks in lineages of rice, *Brachypodium*, and wheat represent ancient centromeres that can be traced to the centromeres of the postulated five ancestral (prior to genome duplication) grass chromosomes A5, A7, A11, A8, and A4 (also called inner circle of cereal genomes, see Bolot et al. 2009) or to their duplicated homologs (Fig. 4). The centromeres of Bd2, W1, and R5 showed perfect homology and trace to the ancestral centromere of A5. The centromere of A7 lineage can be traced to R7 and W2, and persists as an inactivated centromere in Bd1. The centromeres of Bd4, W4, and R11 are homologous and trace their lineage to the ancestral centromere of A11. The centromeres of W7 and R8 showed perfect homology and trace their origin to the ancestral centromere of A8, which is present in Bd3 as an inactivated centromere. Bd5 and R4 are colinear, and their centromeres can be traced to the ancestral centromere of A4.

Centromeres of the duplicated homologs of the ancestral chromosomes conserved in certain rice chromosomes can also be aligned to wheat and *Brachypodium* chromosome centromeres (Fig. 4). W3 and R1 are essentially colinear and, thus, harbor the centromere of A1, a homolog of A5. Both A3 and A10 are homologs of A7. Part of the centromere of W4 may be derived from A3 (R3). Bd3 traces its centromere to A10 (R10). W5 traces its centromere to A12 (R12), a homolog of A11. The ancestral chromosome A4 shares homologs with A6 and A2. Bd1 traces its centromere to A6 (R6). W6 is colinear with R2 and conserves the centromere of A2. There is no evidence for centromere homologies among the duplicated sets of chromosomes, except for the recent report of homology between rice *Cen8* and *Cen9* (Yan et al. 2008).

The first discovery of active genes in rice centromere *Cen8* was surprising and led to the idea that it was a young centromere (Nagaki et al. 2004). However, the conservation of C-gene syntenic blocks among several sets of homologous centromeres of rice, wheat, and *Brachypodium* indicates that active genes can persist in ancient centromeres with more than 40 million years of shared evolutionary history (Paterson et al. 2004). Thus, out of 16 genes in Bd3, nine (64%) were located in the rice *Cen8*, all except one in the same conserved order including six active genes embedded within the 750-kb CENH3 binding domain (Fig. 3; Nagaki et al. 2004; Yan et al. 2005). Many of the same genes also were mapped to the W7 pericentromeric region (Qi et al. 2009). However, ancient centromeres can undergo local structural rearrangements and rapid turnover of centromeric DNA sequences such as satellites and retrotransposons in each lineage as revealed in this study, and even in different species of a genus as has been demonstrated for *Cen8* in different species of rice (Lee et al. 2005; Yan et al. 2008; Gao et al. 2009).

Centromere-associated, syntenic gene blocks of wheat and rice are conserved and associated with centromere-rich repeated arrays in *Brachypodium*. Annotation of a Bd3 BAC-contig sequence homologous to the rice *Cen8* and the W7 pericentromeres indicated that the first 82,594 bp of the 286,663-bp contig sequence lacked genes. In addition, 20,061-bp sequences upstream to the Bd3 contig also did not contain any genes. The 103,655-bp sequences are rich in LTR-*gypsy* retrotransposons. However, no intact or full-length LTR retrotransposons were found in this region, indicating that these are ancient centromeric retrotransposons. We also did not find any sequence similarity in the BAC contig to known centromere retrotransposons conserved in other cereals. These retrotransposon sequences are highly abundant in the Bd centromeres because three Bd3-BAC clones harboring these repeat sequences preferentially hybridized to all Bd centromeres, whereas another two BAC clones in this contig with none or a few LTR-*gypsy* retrotransposons gave FISH signal proximal to the centromere of Bd3 (Table 2, Figs. 1 and 2i and j). Similarly, most of the Bd-BAC clones, located at inactive centromeres in *Brachypodium* chromosomes detected in this study, also competitively hybridized to all Bd centromeres; no FISH signals were observed at the positions of their origin. The results indicated a dramatic accumulation of a large block of LTR/*gypsy*- retrotransposons around active centromeres of *Brachypodium* chromosomes and their loss in inactive centromeric regions.

In contrast, the right border of the Bd3 BAC contig has 16 genes, including six genes syntenic to rice *Cen8* active genes in the CENH3 binding domain (density of 1 gene per 11 kb; Fig. 3b). Nagaki et al. (2004) reported that there were 38 kb of rice CentO repeats in the CENH3 binding domain of *Cen8*. The region with colinear genes between *Brachypodium* and rice *Cen8* in the Bd3 contig contained 81% of tandem repeats detected in the contig. However, none was syntenic to either the rice CentO or BdCENT repeats. Rapid turnover of centromere-specific satellite repeats may also be associated with centromere inactivation because centromeric satellite arrays are the dominant component of most functional centromeres studied so far (Wevrick and Willard 1989; Maluszynska and Heslop-Harrison 1991; Round et al. 1997; Ananiev et al. 1998; Dong et al. 1998; Cheng et al. 2002; Jin et al. 2005). Thus, the centromere inactivation had a major impact on the loss of centromere retrotransposons and turnover of centromere-specific satellites during Bd chromosome evolution.

Acknowledgments We thank Drs. Justin Faris and Brady Vick for critical review of the manuscript, W. John Raupp for editorial assistance, and Duane L. Wilson for excellent technical help. This research was supported by a special USDA-NIFA grant to the Wheat Genetic and Genomic Resources Center. This paper is contribution

number 10-282-J from the Kansas Agricultural Experiment Station, Kansas State University, Manhattan, KS 66506-5502, USA.

References

Ananiev EV, Phillips RL, Rines HW (1998) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA 95:13073–13078

Aragón-Alcaide L, Miller T, Schwatzacher T, Reader S, Moore G (1996) A cereal centromeric sequence. Chromosoma 105:261–268

Bennett MD, Leitch IJ (2005) Nuclear DNA amounts in angiosperms: progress, problems and prospects. Ann Bot 95:45–90

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

Birchler J, Gao Z, Han F (2009) A tale of two centromeres—diversity of structure but conservation of function in plants and animals. Funct Integr Genomics 9:7–13

Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J (2009) The 'inner circle' of the cereal genomes. Curr Opin Plant Biol 12:119–125

Bossolini E, Wicker T, Knobel PA, Keller B (2007) Comparison of orthologous loci from small grass genomes *Brachypodium* and rice: implications for wheat genomics and grass genome annotation. Plant J 49:704–717

Catalan P, Olmstead RG (2000) Phylogenetic reconstruction of the genus *Brachypodium* P-Beauv. (Poaceae) from combined sequences of chloroplast *ndhF* gene and nuclear ITS. Plant Syst Evol 220:1–19

Catalan P, Shi Y, Armstrong L, Draper J, Stace CA (1995) Molecular phylogeny of the grass genus *Brachypodium* P-Beauv based on RFLP and RAPD analysis. Bot J Linnear Soc 117:263–280

Catalan P, Kellogg EA, Olmstead RG (1997) Phylogeny of Poaceae subfamily Pooideae based on chloroplast ndhF gene sequences. Mol Phylogenet Evol 8:150–166

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

Dong F, Jacson MT, SA WGL, Ronald PC, Jinag J (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA 95:8135–8140

Draper J, Mur LAJ, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge APM (2001) *Brachypodium distachyon.* A new model system for functional genomics in grasses. Plant Physiol 127:1539–1555

Faris JD, Zhang Z, Fellers JP, Gill BS (2008) Micro-colinearity between rice, *Brachypodium*, and *Triticum monococcum* at the wheat domestication locus *Q*. Funct Integr Genomics 8:149–164

Francki MG (2001) Identification of Bilby, a diverged centromeric Ty1-copia retrotransposon family from cereal rye (*Secale cereale* L.). Genome 44:266–274

Francki MG, Berzonsky WA, Ohm HM, Anderson JM (2002) Physical location of a HSP70 gene homologue on the centromere of chromosome 1B wheat (*Triticum aestivum* L.). Theor Appl Genet 104:184–191

Gao D, Gill N, Kim HR, Walling JG, Zhang W, Fan C, Yu Y, Ma J, San Miguel P, Jiang N et al (2009) A lineage-specific centromere retrotransposon in *Oryza brachyantha*. Plant J 60:820–831

Garvin DF, Gu YQ, Hasterok R, Hazen SP, Jenkins G, Mockler TC, Mur LAJ, Vogel JP (2008) Development of genetic and genomic research resources for *Brachypodium distachyon*, a new model system for grass crop research. Crop Sci 48:S69–S84

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *Japonica*). Science 296:92–100

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749–752

Hall SE, Kettler G, Preuss D (2003) Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. Genome Res 13:195–205

Han Y, Zhang Z, Liu C, Liu J, Huang S, Jiang J, Jin W (2009) Centromere repositioning in cucurbit species: implication of the genomic impact from centromere activation and inactivation. Proc Natl Acad Sci USA 106:14937–14941

Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G (2006) Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence in situ hybridization. Genetics 173:349–362

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

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DAN. Science 203:1098–1102

Hudakova S, Michalek W, Presting GG, Ten Hoopen R, Dos Santos K, Jasencakova Z, Schubert I (2001) Sequence organization of barley centromeres. Nucleic Acids Res 29:5029–5035

Huo N, Gu YQ, Lazo GR, Vogel JP, Coleman-Derr D, Luo MC, Thilmony R, Garvin DF, Anderson OD (2006) Construction and characterization of two BAC libraries from *Brachypodium distachyon*, a new model for grass genomics. Genome 49:1099–1108

Huo N, Lazo GR, Vogel JP, You FM, Ma Y, Hayden DM, Coleman-Derr D, Hill TA, Dvorak J, Anderson OD et al (2008) The nuclear genome of *Brachypodium distachyon*: analysis of BAC end sequences. Funct Integr Genomics 8:135–147

Huo N, Vogel JP, Lazo GR, You FM, Ma Y, McMahon S, Dvorak J, Anderson OD, Luo M-C, Gu YQ (2009) Structural characterization of *Brachypodium* genome and its syntenic relationship with rice and wheat. Plant Mol Biol 70:47–61

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Jiang J, Nasuda S, Dong F, Scherrer CW, Woo SS, Wing RA, Gill BS, Ward DC (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc Natl Acad Sci USA 93:14210–14213

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

Jin WW, Lamb JC, Vega JM, Dawe RK, Birchler JA, Jiang J (2005) Molecular and functional dissection of the maize B chromosome centromere. Plant Cell 17:1412–1423

Kawabe A, Hansson B, Hagenblad J, Forrest A, Charlesworth D (2006) Centromere locations and associated chromosome rearrangements in *Arabidopsis lyrata* and *A. thaliana*. Genetics 173:1613–1619

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet 33:479–532

Lee HR, Zhang W, Langdon T, Jin WW, 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

Liu Z, Yue W, Li D, Wang RR-C, Kong X, Lu K, Wang G, Dong Y, Jin W, Zhang X (2008) Structure and dynamics of retrotranspo-

sons at wheat centromeres and pericentromeres. Chromosoma 117:445–456

Luo MC, Deal KR, Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ et al (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. Proc Natl Acad Sci USA 106:15780–15785

Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci USA 103:5224–5229

Maluszynska J, Heslop-Harrison JS (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. Plant J 1:159–166

McCarthy EM, McDonald JF (2003) LTR STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics 19:362–367

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 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

Opanowicz M, Vain P, Draper J, Parker D, Doonan JH (2008) *Brachypodium distachyon*: making hay with a wild grass. Trends Plant Sci 13:172–177

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908

Paterson AH, Bower JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A Ty3-*gypsy* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16:721–728

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvorák J, Linkiewicz AM, Ratnasiri A et al (2004) A chromosome bin map of 16, 000 expressed sequence tag loci and distribution of genes among the three genomes of polyploidy wheat. Genetics 168:701–712

Qi LL, Friebe B, Gill BS (2006) Complex genome rearrangements reveal evolutionary dynamics of pericentromeric regions in the Triticeae. Genome 49:1628–1639

Qi LL, Friebe B, Zhang P, Gill BS (2009) A molecular-cytogenetic method for locating genes to pericentromeric regions facilitates a genome-wide comparison of synteny between the centromeric regions of wheat and rice. Genetics 183:1235–1247

Round EK, Flowers SK, Richards EJ (1997) *Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure. Genome Res 91:1007–1019

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, Calcagno T, Cooke R, Delseny M, Feuillet C (2008) Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell 20:11–24

Sandhu D, Champoux JA, Bondareva SN, Gill KS (2001) Identification and physical localization of useful genes and markers to a major gene-rich region on wheat 1 S chromosomes. Genetics 157:1735–1747

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Sharma A, Presting GG (2008) Centromeric retrotransposon lineages predate the maize/rice divergence and differ in abundance and activity. Mol Genet Genomics 279:133–147

Shi Y, Draper J, Stace C (1993) Ribosomal DNA variation and its phylogenetic implications in the genus *Brachypodium* (Poaceae). Plant Syst Evol 188:125–138

Stein N, Prasad M, Scholz U, Thiel T, Zhang HG, Wolf M, Kota R, Varshney RK, Perovic D et al (2007) A 1, 000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet 114:823–839

The International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

Vogel JP, Gu YQ, Twigg P, Lazo GR, Laudencia-Chingcuanco D, Hayden DM, Donze TJ, Vivian LA, Stamova B, Coleman-Derr D (2006) EST sequencing and phylogenetic analysis of the model grass *Brachypodium distachyon*. Theor Appl Genet 113:186–195

Wevrick R, Willard HF (1989) Long-range organization of tandem arrays of a satellite DNA at the centromeres of human chromosomes: high frequency array-length polymorphism and meiotic stability. Proc Natl Acad Sci USA 86:9394–9398

Wolfgruber TK, Sharma A, Schneider KL, Albert PS, Koo DH, Shi J, Gao Z, Han F, Lee H, Xu R et al (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. PLoS Genet 5:e100743

Wong LH, Choo KHA (2004) Evolutionary dynamics of transposable elements at the centromere. Trends Genet 20:611–616

Wu J, Yamagata H, Hayashi-Tsugane M, Hijishita S, Fujisawa M, Shibata M, Ito Y, Nakamura M, Sakaguchi M, Yoshihara R et al (2004) Composition and structure of the centromeric region of rice chromosome 8. Plant Cell 16:967–976

Wu J, Fujisawa M, Tian Z, Yamagata H, Kamiya K, Shibata M, Hosokawa S, Ito Y, Hamada M, Katagir S et al (2009) Comparative analysis of complete orthologous centromeres from two subspecies of rice reveals rapid variation of centromere organization and structure. Plant J 60:805–819

Yan HH, Jin W, Nagaki K, Tian S, Ouyang S, Buell CR, Talbert PB, Henikoff S, Jiang J (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, Ouyang S, Tian JW, SL LuC, Venu RC, Wang GL, Green PJ 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

Yan HH, Talbert PB, Lee HR, Jett J, Henikoff S, Chen F, Jiang J (2008) Intergenic locations of rice centromeric chromatin. PLoS Biol 6:e286

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X et al (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296:79–92

Yu J, Wang J, Lin W, Li S, Li H, Zhou J, Ni P, Dong W, Hu S, Zeng C et al (2005) The genomes of *Oryza sativa*: a history of duplications. PLoS Biol 3:e38

Zhang P, Li WL, Fellers J, Friebe B, Gill BS (2004a) BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. Chromosoma 112:288–299

Zhang Y, Huang Y, Zhang L, Li Y, Lu T, Lu Y, Feng Q, Zhao Q, Cheng Z, Xue Y et al (2004b) Structural features of the rice chromosome 4 centromere. Nucleic Acids Res 32:2023–2030