

This is the author's final, peer-reviewed manuscript as accepted for publication. The publisher-formatted version may be available through the publisher's web site or your institution's library.

Sequence organization and evolutionary dynamics of Brachypodium-specific centromere retrotransposons

L. L. Qi, J. J. Wu, B. Friebe, C. Qian, Y. Q. Gu, D. L. Fu, and B. S. Gill

How to cite this manuscript

If you make reference to this version of the manuscript, use the following information:

Qi, L. L., Wu, J. J., Friebe, B., Qian, C., Gu, Y. Q., Fu, D. L., & Gill, B. S. (2013). Sequence organization and evolutionary dynamics of Brachypodium-specific centromere retrotransposons. Retrieved from <http://krex.ksu.edu>

Published Version Information

Citation: Qi, L. L., Wu, J. J., Friebe, B., Qian, C., Gu, Y. Q., Fu, D. L., & Gill, B. S. (2013). Sequence organization and evolutionary dynamics of Brachypodium-specific centromere retrotransposons. *Chromosome Research*, 21(5), 507-521.

Copyright: © Springer Science+Business Media Dordrecht (Outside the USA) 2013

Digital Object Identifier (DOI): doi:10.1007/s10577-013-9378-4

Publisher's Link: <http://link.springer.com/article/10.1007/s10577-013-9378-4>

This item was retrieved from the K-State Research Exchange (K-REx), the institutional repository of Kansas State University. K-REx is available at <http://krex.ksu.edu>

1 Sequence organization and evolutionary dynamics of *Brachypodium*-specific centromere
2 retrotransposons

3

4 L.L. Qi^{1*†}, J.J Wu^{2*}, B. Friebe³, C. Qian³, Y.Q. Gu⁴, D.L. Fu², and B.S. Gill³

5

6 ¹USDA-ARS, Northern Crop Science Laboratory, 1605 Albrecht Blvd N, Fargo, ND
7 58102-2765 USA

8 ²State Key laboratory of Crop Biology, College of Agronomy, Shandong Agricultural
9 University, 61 Daizong Street, Taian, Shandong, 271018, P. R. China

10 ³Wheat Genetic Resources Center, Department of Plant Pathology, Kansas State
11 University, Manhattan, KS 66506-5502, USA

12 ⁴USDA-ARS, Genomics and Gene Discovery Research Unit, Western Regional Research
13 Center, 800 Buchanan Street, Albany, CA, 94710 USA

14 *These authors contributed equally to this work

15 †For correspondence (phone +1 701 239 1351; fax +1 701 239 1346; email

16 lili.qi@ars.usda.gov)

17

1 Abstract

2

3 *Brachypodium distachyon* is a wild annual grass belonging to the Pooideae, more closely
4 related to wheat, barley, and forage grasses than rice and maize. As an experimental
5 model, the completed genome sequence of *B. distachyon* provides a unique opportunity
6 to study centromere evolution during the speciation of grasses. Centromeric satellite
7 sequences have been identified in *B. distachyon*, but little is known about centromeric
8 retrotransposons in this species. In the present study, BAC-fluorescence *in situ*
9 hybridization was conducted in maize, rice, barley, wheat, and rye using *B. distachyon*
10 (Bd) centromere-specific BAC clones. Eight Bd centromeric BAC clones gave no
11 detectable FISH signals on the chromosomes of rice and maize, and three of them also
12 did not yield any FISH signals in barley, wheat, and rye. In addition, four of five
13 Triticeae centromeric BAC clones did not hybridize to the *B. distachyon* centromeres,
14 implying certain unique features of *Brachypodium* centromeres. Analysis of
15 *Brachypodium* centromeric BAC sequences identified a long terminal repeat (LTR)-
16 centromere retrotransposon of *B. distachyon* (*CRBd1*). This element was found in high
17 copy number accounting for 1.6% of the *B. distachyon* genome, and is enriched in
18 *Brachypodium* centromeric regions. *CRBd1* accumulated in active centromeres, but was
19 lost from inactive ones. The LTR of *CRBd1* appears to be specific to *B. distachyon*
20 centromeres. These results reveal different evolutionary events of this retrotransposon
21 family across grass species.

22

23 Introduction

1 As chromosome landmarks, centromeres are responsible for kinetochore assembly that
2 links chromosome to microtubule spindle, and thereby enabling the faithful segregation
3 of sister chromatids during cell division. Extensive tracts of tandem repeats (centromeric
4 satellites) interrupted by various retrotransposons are common structural features of
5 centromeres (Copenhaver et al. 1999; Kumekawa et al. 1999, 2001; Jiang et al. 2003;
6 Zhang Y et al. 2004; Lamb et al. 2008). Satellite DNA and centromeric retrotransposons
7 (CR) are the most abundant DNA elements found in plant centromeres and are associated
8 with CENH3, a centromere-specific histone H3 present in nucleosomes of active
9 centromeres (Jiang et al. 1996; Miller et al. 1998; Presting et al. 1998; Cheng et al. 2002;
10 Zhong et al. 2002; Nagaki et al. 2003b, 2004).

11

12 Centromeric satellite DNA sequences have been isolated from several plant species,
13 including *Arabidopsis* (Round et al. 1997), maize (Ananiev et al. 1998), sorghum (Miller
14 et al. 1998), rice (Cheng et al. 2002; Zhang Y et al. 2004; Lee et al. 2005), *Medicago*
15 *truncatual* (Kulikova et al. 2004), *Brassica* (Lim et al. 2007), *Brachypodium* (2010), and
16 soybean (Tek et al. 2010). Although the repeat length, ranging in size from 155 bp (rice)
17 to 180 bp (*Arabidopsis*), is similar between taxa, their sequences are largely species-
18 specific and highly divergent even between closely related species (Lee et al. 2005). In
19 contrast to the centromeres of most plant species where functional centromeres are
20 mainly composed of large arrays of centromere satellite repeats and CR elements (Jiang
21 et al. 2003), wheat centromeres lack tandem satellite repeats of megabase size and are
22 dominated by centromeric retrotransposons (Liu et al. 2008; Li et al. 2013).

1 Unlike centromeric satellites, the *CR* family in grass species is highly conserved. Two
2 highly conserved *CR* sequences, *CCSI* and *pSau3A9*, which are parts of Ty3/*gypsy*-type
3 retrotransposons, were first found to localize at the centromeres of most cereal species
4 that have been investigated (Aragon-Alcaide et al.1996; Jiang et al 1996, Miller et al.
5 1998; Presting et al. 1998). *CRR* (*CR* of rice) and *CRM* (*CR* of maize) are the most
6 intensively studied *CR* elements among plant species (Dong et al. 1998; Presting et al.
7 1998; Cheng et al. 2002; Zhong et al. 2002; Nagaki et al. 2003a, 2005). Rice *CRR1* is
8 homologous to maize *CRM3*, *CRR2* to *CRM2*, *CRR3* to *CRM1*, and *CRR4* to *CRM4*,
9 which pre-date the divergence of maize and rice (Sharma and Presting 2008). Two
10 putative *CR* families of soybean were also grouped to *CRR* and *CRM* lineage (Du et al.
11 2010) and the *CR* elements, *Beetle1* and *Beetle 2*, found in beet are highly similar to the
12 *CRs* of rice, maize, and Barley (Weber and Schmidt 2009). The *CR* elements isolated
13 from barley and wheat showed cross hybridization among cereal species (Hudakova et al.
14 2001; Zhang P et al. 2004). As few exceptions to the general *CR* conservation of grasses,
15 species-specific *CR* element was reported in rye (Francki 2001) and wild rice (Gao et al.
16 2009). A rye-specific *CR*, *Bilby* that is a Ty1-*copia* retrotransposon-like element, is
17 highly divergent from other known cereal *CR* elements, and a lineage-specific *CR*
18 element was identified in *Oryza brachyantha*.
19
20 *Brachypodium distachyon* (hereafter referred to as *Brachypodium*) is a wild annual grass
21 belonging to the Pooideae, more closely related to wheat, barley, and forage grasses than
22 rice and maize. As an experimental model, the completed genome sequence of
23 *Brachypodium* provides an important reference for grass biology and centromere studies

1 (The International Brachypodium Initiative, 2010). A 156 bp *Brachypodium* centromeric
2 repeat (Bd_CENT) was identified and is present on all the *Brachypodium* centromeres
3 (The International Brachypodium Initiative, 2010). The only completely assembled
4 centromere is 45 kb long on chromosome Bd5 and is composed of two Bd_CENT arrays
5 occasionally interspersed with large blocks of unknown LTR retrotransposons (The
6 International Brachypodium Initiative, 2010). Previous studies indicated that the gene
7 sequences in the centromeric and pericentromeric regions from rice and wheat were
8 conserved with those in the centromeric/pericentromeric regions of *Brachypodium*,
9 indicating that these genes pre-existed in the centromere regions before the divergence of
10 the grass species that occurred 50-70 MYA (Bolot et al. 2009; Qi et al. 2010). However,
11 54 genes found within 300 kb of all five *Brachypodium* centromeres were non-collinear
12 with rice and sorghum, indicating some unique features of *Brachypodium* centromeres
13 after it diverged from rice and wheat (The International Brachypodium Initiative, 2010).
14 In order to study the evolution of *Brachypodium* centromeres, we conducted BAC-
15 fluorescence *in situ* hybridization in maize, rice, barley, wheat and rye using 19
16 *Brachypodium* centromere-specific BAC clones, and annotated in detail four of these
17 BAC clones. The results demonstrate that *Brachypodium* CR elements are highly
18 divergent from those of other grass species.

19

20 Materials and Methods

21

22 Materials

23

1 Seeds of *B. distachyon*, an inbred, diploid line Bd21, were obtained from USDA-ARS,
2 Pacific West Area, Western Regional Research Center, Genomics and Gene Discovery,
3 Albany, CA, USA. Chinese Spring (CS) wheat (*Triticum aestivum* L.), Imperial rye
4 (*Secale cereale* L.), Betzes barley (*Hordeum vulgare* L.) were provided by the Wheat
5 Genetic Resources Center at Kansas State University, KS, USA. Nipponbare rice (*Oryza*
6 *sativa* L.), and B73 maize (*Zea mays* L.) were provided by Drs. Frank White and Harold
7 Trick at the Plant Pathology Department, Kansas State University, KS, USA.

8

9 Methods

10

11 Selection of *Brachypodium* putative centromeric BAC

12

13 Three centromere-specific clones, Hi10, pRCS1, and pAet6-09, were used in the present
14 study. Both Hi10 and pRCS1 are cereal-specific centromeric DNA sequences; Hi10 was
15 isolated from *B. sylvaticum* (Abbo et al. 1995), and pRCS1 was derived from rice (*O.*
16 *sativa* ssp. Indica cv. IR-BB21) (Dong et al. 1998). The clone pAet6-09 was isolated
17 from *Ae. tauschii* bacterial artificial chromosome (BAC) library and hybridized to the
18 centromeres of wheat, barley, rye, and maize, but not to rice (Zhang P et al. 2004). Hi10
19 and pRCS1 were used to screen one high-density filter containing 18,432 clones from
20 *Brachypodium* BAC library (~4.5 × coverage) (Huo et al. 2006). The BAC clones with
21 unambiguous positive hybridization signals were selected, digested with *HindIII*, and
22 hybridized again to the three clones, Hi10, pRCS1, and pAet6-09. The putative
23 centromeric BAC clones were selected as probes for further BAC-fluorescence *in situ*

1 hybridization (FISH) experiments (Table 1). Five additional *Brachypodium* BAC clones,
2 which previously gave BAC-FISH signals at the centromeres of *Brachypodium*
3 chromosomes were also used in the present study (Table 1, Qi *et al.* 2010). These BAC
4 clones were anchored by wheat pericentromeric ESTs from homoeologous chromosome
5 groups 3, 4, and 6. The procedure for colony filter hybridization and southern
6 hybridization was described by Qi *et al.* (2009).

7
8 Selection of the centromeric BAC clones from wheat 3B, *Aegilops speltoides*, and
9 *Aegilops tauschii* BAC libraries

10

11 Two wheat centromeric BAC clones, 3B-100-L17 and 3B-40-L07, were obtained by
12 screening a wheat 3B BAC library (Šafář *et al.* 2004) using the pAet6-09 sequence as
13 probe (Qi *et al.* unpublished data, Table 1). These two clones are located on the ordered
14 BACs of contig 796 (Feuillet, personal communication). Later, 3B-100-L17 as a single
15 BAC was placed in the 3B centromere along with 12 sequenced contigs in 3B
16 chromosome (Choulet *et al.* 2010). Both clones exclusively hybridized to the centromeres
17 of wheat chromosomes (Fig. 1a). Two *Ae. speltoides* centromeric BAC clones, 21E12
18 and 256 K19, and one *Ae. tauschii* centromeric BAC clone HD008H01 were identified
19 previously by Qi *et al.* (2009) (Table 1).

20

21 BAC-fluorescence *in situ* hybridization (BAC-FISH)

22

1 Mitotic chromosome spreads for BAC-FISH were prepared from the root tips of
2 *Brachypodium* Bd21, CS wheat, Imperial rye, Betzes barley, Nipponbare rice, and B73
3 maize as described by Qi et al. (2010). BAC DNA was isolated using a Qiagen Plasmid
4 Midi Kit following the manufacture's instruction (Qiagen Valencia, Calif.). One
5 microgram of BAC DNA was labeled with fluorescein-12-dUTP (Enzo Life Science Inc,
6 Farmingdale, NY) using nick translation. The BAC-FISH was performed on metaphase
7 chromosomes as previously described (Qi et al. 2010). Slides were analyzed with an
8 epifluorescence Zeiss Axioplan 2 microscope. Images were captured using a SPOT 2.1
9 CCD (charge-coupled device) camera (Diagnostic Instruments) and processed with
10 Photoshop v5.5 (Adobe Systems).

11

12 BAC sequence annotation

13

14 Four BAC clones, DB069J23, DB088O14, DB042E22 and DH017G05, were end
15 sequenced and then were anchored onto the *Brachypodium* chromosomes. A 300-kb
16 continuous stretch of sequence extending from one BAC-end was used for annotation
17 analysis. Self alignment of each BAC was performed using NCBI bl2seq BLAST tool to
18 generate a first glance of its repetitive nature. The RepeatScout (Price et al. 2005) was
19 also used to identify *de novo* repeats. To find common sequences among these BACs,
20 they were also aligned with each other. Transposable elements were identified by a
21 combination of BLAST searches against the GenBank nonredundant database and the
22 Triticeae Repeat Sequence Database (*TREP*, <http://wheat.pw.usda.gov/ITMI/Repeats/>).
23 LTR-FINDER (Zhao and Wang, 2007) was used to predict full-length LTR

1 retrotransposons with tRNA database of *Brachypodium*. The insertion time of
2 retrotransposons was calculated according to Ma and Bennetzen (2004). The frequency
3 and distribution of the repeat elements along the chromosomes was analyzed by
4 searching the *Brachypodium* genome assembly (<http://www.brachypodium.org/>) with
5 NCBI local BLAST tool kit version 2.2.11, and the e-value cut-off was set to 1e-10. The
6 NCBI database, PlantGDB, CerealsDB, barley (webblast.ipk-gatersleben.de/barley/) and
7 rice (<http://rice.plantbiology.msu.edu/>) whole genome sequences were BLASTN searched
8 to identify conserved sequences of *Brachypodium* repeats.

9

10 Results

11

12 Identification of *Brachypodium*-specific centromeric BACs

13

14 The *B. distachyon* BAC library (~4.5 × coverage) was probed with Hi10 and
15 pRCS1 clones. Of the 38 unambiguous positive BAC clones, 13 were selected by probe
16 Hi10 and 25 by pRCS1 (Table S1). The BAC clones were digested with restriction
17 enzyme *Hind*III and hybridized to the three centromeric DNA sequences, Hi10, pRCS1,
18 and pAet6-09. Positive Southern hybridization signals and ladder patterns were detected.
19 Average numbers of BAC fragments hybridizing to three clones were 6.7 (range from 0
20 to 15) for Hi10, 7.0 (1 to 16) for pAet6-09, and 2.0 (0 to 4) for pRCS1 (Table S1).
21 Fourteen BAC clones that gave intense hybridization signals were selected for further
22 BAC-FISH with chromosome complements of *Brachypodium* and rye (Table S1).

23

1 FISH results of the selected BAC clones to *Brachypodium* chromosomes indicated that
2 14 BAC clones exclusively hybridized to the centromeric regions of all *Brachypodium*
3 chromosomes with very strong FISH signals (Table 2, Fig. 2d). Subsequently, these BAC
4 clones along with other five *Brachypodium* BAC clones, DH017G05, DH039C01,
5 DB069J23, DB042E22, and DB088O14, which previously showed FISH signals on the
6 centromeric regions of all the five *Brachypodium* chromosomes (Qi et al. 2010), were
7 FISH mapped on rye chromosomes. The FISH on rye chromosomes showed variable
8 signal intensities exclusively at the primary constructions (Table 2, Figs. 2b and 3c).
9 Based on signal intensity, these 19 BACs were divided into four groups: group I with
10 three BACs showed strong signals (Fig. 2b), group II with six BACs gave faint signals
11 (Fig. 3c), group III with seven BACs showed the very weak signals on the rye
12 centromeres when image was exposed longer than usually required for detecting the
13 corresponding signals on *Brachypodium* centromeres, and group IV with three BACs
14 showed no FISH signals (Fig. 1d). To identify *Brachypodium*-specific centromeric BAC
15 clones, eight BAC clones were selected, all of which except DH007B23, gave very strong
16 signals on *Brachypodium* centromeric regions, but yielded variable signal intensities on
17 those of rye (Table 2), and were used to hybridize to chromosome complements of wheat,
18 barley, rice, and maize.

19

20 Surprisingly, no detectable FISH signal of eight selected BAC clones was observed on
21 the chromosomes of rice and maize, indicating that canonical sequences of
22 *Brachypodium* CR elements appear to have disappeared in these two species (Figs. 2e, 2f,
23 3d, and 3f). Out of eight BAC clones, three: DH017G05, DB042E22, and DB088O14,

1 with a similar result on rye, also did not hybridize to the chromosomes of wheat and
2 barley, indicating that the centromere-specific sequences present in these BACs are
3 sufficiently conserved only in *Brachypodium* (Table 2, Fig. 1c and d). The remaining five
4 BAC clones showed variation in the intensity of the FISH signals in wheat, rye, and
5 barley. BAC DH021M4 had very strong FISH signals on the centromeres of
6 *Brachypodium* chromosomes, as well as in rye, wheat and barley. The FISH signal
7 intensity of BAC clones: DH010C12 and DH029E4, was similar in rye and wheat, but
8 lower in barley (Fig. 2a-c). BAC clones, DH008A23 and DH007B23, gave a weak
9 centromeric FISH signals in rye and wheat, but no signals in barley (Fig. 3a-c).
10 BAC-FISH of Triticeae centromeric BAC clones to *Brachypodium* chromosomes

11

12 Five centromeric BAC clones from wheat, *Ae. speltoides*, and *Ae. tauschii* were analyzed
13 for their hybridization to the chromosomes of *Brachypodium*. Only the BAC clone, 3B-
14 40-L07 derived from the wheat 3B BAC library, showed FISH signals on the centromeres
15 of *Brachypodium* chromosomes. Another 3B BAC clone, 3B-100-L17 residing at the
16 centromere of 3B chromosome, did not give any FISH signals on the *Brachypodium*
17 chromosomes (Fig. 1b, Table 3). Neither the two BAC clones from *Ae. speltoides* nor the
18 one from *Ae. tauschii* produced any FISH signals in *Brachypodium*, indicating that the
19 centromeric repeats in Triticeae have diverged from those in *Brachypodium*

20

21 Sequence organization of *Brachypodium*-specific centromeric BAC clones

22

23 *Common repeats in the four Brachypodium BACs*

1
2 Sequence annotation was performed in four *Brachypodium* BAC clones, DH017G05,
3 DB069J23, DB042E22, and DB088O14. Because these BACs produced strong signals
4 only on *Brachypodium* centromeres in FISH experiments, they were all tested for the
5 presence of the 156 bp centromeric satellite repeat Bd_CENT (The International
6 Brachypodium Initiative, 2010) and found to be negative. The self alignments revealed
7 that DB069J23 was highly repetitive, and multiple copies of repeats were scattered in a
8 225 kb region (Fig. S1). Further comparison indicated that this repetitive region in
9 DB069J23 was also present in the other three BACs with variable copies. Using
10 RepeatScout program, a total of 26 repeat elements (>4 copies) were identified in
11 DB069J23, and nine of them were repeated more than 10 times. The most abundant
12 repeats could be assembled into two contigs with >80% sequence similarity, indicating
13 they belong to two repeat elements. Based on the Blastn searches against DB069J23 and
14 the *Brachypodium* whole genome sequence, these two repetitive contigs were represented
15 by two fragments in DB069J23, designated as RM-1 (DB069J23:107658-109770; 2113
16 bp) and RM-2 (DB069J23:271259-271757; 499 bp). In DB069J23, RM-1 fragment had
17 46 copies with a total accumulative length of 49.3 kb and RM-2 had 24 copies with an
18 accumulative length of 9.6 kb (e-value < 1e-10). The conserved sequences of RM-1 in
19 DB088O14, DB042E22 and DH017G05 were 12.8, 2.4 and 9.8 kb, respectively. As for
20 RM-2, that was 3.7, 1.5 and 2.4 kb, respectively.

21

22 *RM-1 and RM-2 are a part of a single LTR specific to Brachypodium*

1 Detailed analysis of the *Brachypodium* genomic regions containing RM-1 and RM-2
2 revealed that these two repeats were parts of the LTR of one *Gypsy* retrotransposon found
3 in BAC DB088O14, designated as *CRBdl* (Centromeric Retrotransposon of *B*.
4 *distachyon*, GeneBank # KF040483) (Fig. 4a). RM-1 and RM-2 were located
5 immediately at the 3' and 5' end of the LTR, respectively. The complete LTR was about
6 3.3 kb in length, and between the RM-1 and RM-2 was a region with high GC content
7 (~70%). Eight full length copies of *CRBdl* were identified in the *Brachypodium* genome,
8 with size ranging from 12.5 kb to 12.8 kb (Supplementary file 1), and five of them were
9 located less than 3 Mb from the centromeres or chromosome fusion points (Fig. 5, Table
10 S2). The insertion time of the eight full length *CRBdl* was estimated to be in the range of
11 0.01 to 1.34 million years ago (Mya). Target site duplications (TSD) were found in 7 of
12 them. In the internal region (~6.2 kb) between two LTRs, there were one Zinc knuckle
13 domain (pfam: zf-CCHC) and one chromatin organization modifier domain (pfam:
14 Chromo), besides feature proteins present in a typical LTR retrotransposable element
15 such as retrotransposase (Fig. 4a). Using the chromodomain identified in *CRBdl* as query
16 sequence, a total of 420 copies (e-value < 1e-5) were identified in whole *Brachypodium*
17 genome. The distribution patterns of chromodomain are well consistent with that of
18 *CRBdl* which is enriched in centromere regions (Fig. 5).

19

20 No full length element of *CRBdl* was identified in the four *Brachypodium* BACs. In BAC
21 DB088O14, two truncated and one fragmented elements, and one partial LTR and one
22 solo LTR were clustered in a 65.7 kb region in a different retrotransposon (Fig. 4b). The
23 *CRBdl* homologous sequences occupied a total of 44.6 kb (67.9%) of this region. In

1 comparison, *CRBd1* element was more abundant in DB069J23 than in DB088O14 (Table
2 5), and the total length of homologous region was 101.2 kb. More copies of the LTR than
3 the internal region of *CRBd1* were present in DB069J23 (Fig. 4c), and 6 solo LTRs with
4 TSD were identified.

5
6 By searching the *Brachypodium* genome assembly with e-value of 1e-10, *CRBd1* totally
7 occupied 4.41 Mb in length and constituted 1.6% of the whole genome. Consistent with
8 the distribution pattern in DB069J23 (Fig.4c), the LTR of the *CRBd1* was much more
9 abundant than the internal region, and many solo LTRs could be identified in the
10 *Brachypodium* genome (data not shown). The LTR of *CRBd1* was obviously enriched in
11 the centromeric regions or Bd_CENT containing regions (Fig. 5). The top three most
12 abundant regions included the centromere of Bd chromosome 1, and two chromosome
13 fusion points (The International Brachypodium Initiative, 2010) on the long arm of Bd
14 chromosome 2 and the short arm of Bd chromosome 3. The total length of the sequences
15 homologous to the LTR is 2.76 Mb in the *Brachypodium* genome.

16
17 Based on the sequence comparison to other grass species, the LTR of *CRBd1* was
18 specific to *Brachypodium*. Only few sequences were found in wheat and barley with
19 limited conservation to small sections of the LTR. Given the genome coverage of the
20 current Triticeae datasets, the copy number of the homologous sequence should be very
21 low in wheat or barley. No conserved LTR sequence was found in rice, maize, sorghum
22 and rye. However, in these species, retrotransposons could be found that had about 70%
23 DNA sequence identity to the internal region of *CRBd1*, and the retrotransposons also

1 had the conserved zf-CCHC and Chromo domains. Furthermore, the boundary sequences
2 (10-15bp) of the LTRs were conserved among the cereal species (Fig. 6).

3 *Other retrotransposons in the four Brachypodium BACs*

4

5 Besides *CRBd1*, another four retrotransposons were identified in at least one BAC (Table
6 4). *CRBd2* (GeneBank # KF040484, Fig. 4b) was identified in DB088O14 as a full length
7 *Gypsy* retrotransposon that showed 73% DNA sequence similarity to the rice centromeric
8 retrotransposon CRR3 (GenBank # DQ458292). One full length *CRBd2* and one solo
9 LTR were found in DB069J23 and DH017G05, respectively. The distribution of *CRBd2*
10 was also associated with centromeric regions or Bd_CENT containing regions (data not
11 shown). Its low content in the *Brachypodium* genome (263.7kb, 0.097%) indicates that
12 *CRBd2* was not highly repetitive (Table 4). Other three common retrotransposons which
13 we term 'retrotransposon element of *B. distachyon*' (REBd) including two *Gypsy*
14 retrotransposons, *REBd1* (GeneBank # KF0404850) and *REBd2* (GeneBank #
15 KF040486), and one *Copia* element *REBd3* (GeneBank # KF040487), were found in
16 DB088O14, DB069J23, and DH017G05, respectively (Table 4). These retrotransposons
17 were randomly distributed along the chromosomes without any obvious association with
18 centromeric regions.

19

20 In DB088O14, we observed the amplification of *CRBd1* in the internal region of *REBd1*
21 element (Fig. 4b). After at least two rounds of insertion and deletion, one *CRBd1* cluster
22 formed contained three truncated *CRBd1* and two solo LTRs. Based on a substitution rate
23 of 1.3×10^{-8} per site per year, the *REBd1* containing the *CRBd1* cluster was originally

1 inserted in between the genes Bradi3g44470 and Bradi3g44480 about 2.71 million years
2 ago (MYA). Another *REBd1* retrotransposon at the 3' end of BAC DB088O14 was
3 inserted 3.27 MYA. In comparison, the *REBd3* was a young retrotransposon inserted 0.23
4 MYA and another homolog *REBd3* in DB042E22 was inserted 0.03 MYA with only one
5 nucleotide substitution between the two LTRs (1328 bp).

6

7 Discussion

8

9 *Brachypodium* centromeres, similar to cereal centromeres (rice, maize, and sorghum),
10 mainly consist of centromeric satellite sequences and retrotransposons (The International
11 Brachypodium Initiative 2010; Wen et al. 2012). Centromeric satellite sequences have
12 evolved and diverged rapidly and are largely species-specific, whereas centromere
13 retrotransposons (CR) appear to evolve more slowly (Round et al. 1997; Ananiev et al.
14 1998; Copenhaver et al. 1999; Henikoff et al. 2001; Cheng et al. 2002; Jin et al. 2004,
15 2005; Hall et al. 2003; Lee et al. 2005; Tek et al 2010). In the cereal species, *CRRs* in
16 rice, *CRMs* in maize, *CRWs* and *Quinta* in wheat, and *Cereba* in barley are highly
17 conserved across related genomes and over long evolutionary periods (Dong et al. 1998;
18 Miller et al. 1998; Presting et al. 1998; Hudakova et al. 2001; Zhang P et al. 2004; Liu et
19 al. 2008; Sharma and Presting 2008; Li et al. 2013). However, it was surprising to
20 observe that eight selected *Brachypodium* centromeric-BAC clones did not hybridize to
21 any centromeres of rice and maize. In addition, three of them also did not hybridize to the
22 centromeres of rye, wheat, and barley (Table 2).

23

1 Sequence annotation of the four BACs revealed that two repetitive elements, RM-1 and
2 RM-2 , were abundant in the *Brachypodium* genome, and belong to parts of the LTR of a
3 truncated *gypsy* retrotransposon, *CRBd1*, derived from BAC DB088O14 (Fig. 4). *CRBd1*
4 element was represented in all four BACs, and accounted for 4.4 Mb (1.6%) of the *B.*
5 *distachyon* genome (Table 4). The LTR of *CRBd1* harboring the RM-1 and RM-2 is
6 enriched in *Brachypodium* centromeric regions (Fig. 5), and is appeared to be a
7 *Brachypodium* specific sequence. Although LTRs usually diverge faster than the other
8 parts of the retrotransposons, highly conserved DNA motifs were found in the LTRs of
9 the *CR* elements from rice, maize, and barley (Nagaki et al. 2003a). However, our results
10 indicate that the 3'end LTR of *CRBd1* is significantly diverged in the other grass
11 genomes tested and had undergone rapid amplification in the regions of the currently
12 active centromeres during evolution of *B. distachyon* centromeres.

13

14 Comparative sequence analysis between *Brachypodium*, wheat, rice, and sorghum
15 revealed nested insertions of entire chromosomes into centromeric regions during the
16 evolution of five *Brachypodium* chromosomes from a 12-chromosome ancestor of all
17 grasses (The International Brachypodium Initiative 2010; Qi et al. 2010). Three of four
18 BACs analyzed, DH017G05, DB042E22, and DB088O14, were located at inactive
19 centromeres of Bd chromosomes 2 and 3, and both DH017G05 and DB042E22 are in the
20 fusion points of these chromosomes, indicating that these BAC clones were originally
21 located at the centromeric regions of ancestral chromosomes (Fig. 5). However, all these
22 BAC clones do not contain *Brachypodium* centromere satellite repeat, Bd_CENT. FISH
23 results showed that they landed to the active centromeres of *Brachypodium*, and none

1 yielded FISH signals at their original positions. These results imply that the accumulation
2 of the *CR* elements originally present in these BACs have preferentially occurred in the
3 regions of the currently active centromeres. It is also evident that the *CR* element of
4 *CRBd1* is more abundant in BAC DB069J23, which is located at the active centromere
5 region of Bd chromosome 4 (Fig. 5). The conserved sequence of *CRBd1* were 101.2 kb in
6 length in DB069J23 compared to 44.6 kb in DB088O14, 20.8 kb in DH017G05, and 7.5
7 kb in DB042E22 (Table 4). These results support the hypothesis that redundant
8 centromeres in *Brachypodium* chromosomes became inactive by the loss of centromeric
9 retrotransposons and rapid turnover of centromere-specific satellites (Qi et al. 2010). In
10 other words, the *Brachypodium* active centromeres maintain centromere satellite repeats
11 and accumulate *CR* elements as a result of centromere drive (Ma et al. 2007; Wu et al.
12 2009). Only eight full length retrotransposons of *CRBd1* were found in the *Brachypodium*
13 genome, and many solo-LTR of *CRBd1* are present in the four BACs analyzed and in the
14 *Brachypodium* genome, revealing that *CRBd1* is an ancient centromeric retrotransposon.
15
16 Rice and *Brachypodium* diverged approximately 40-54 MYA, while *Brachypodium* and
17 wheat diverged approximately 30 MYA (The International Brachypodium Initiative
18 2010). Although their genomes vary in size and basic chromosome numbers, gene
19 content and gene order has been largely conserved. The conserved genes were also
20 reported to be present in the centromere regions of rice, wheat, and *Brachypodium*, which
21 share the syntenic blocks among several sets of homologous centromeres (Qi et al. 2010).
22 However, the *Brachypodium* *CR* elements appear to be highly divergent from other grass
23 species, especially from rice and maize. Except three BAC clones mentioned above, five

1 other *Brachypodium* centromeric BAC clones also did not yield any FISH signals in rice
2 and maize. These BAC clones were obtained by screening *Brachypodium* BAC library
3 using Hi10 as probe (Table 2). Hi10 was isolated from *B. sylvaticum*, a species diverged
4 from *B. distachyon* approximately 1.7-2.0 MYA (Buchmann et al. 2012), and contains
5 CCS1 sequence belonging to a cereal centromeric retrotransposon (Abbo et al. 1995;
6 Aragon-Alcaide et al. 1996). Wen et al. (2012) also reported that CCS1 failed to label *B.*
7 *distachyon* centromeres and its homologous sequences are comparatively less abundant in
8 the *B. distachyon* genome. In addition, all five Triticeae centromeric BAC clones used in
9 the present study, except one, 3B-40-L07, did not yield any FISH signals in the
10 *Brachypodium* centromeres (Table 3) (Qi et al. 2009, 2010). Among them, the 3B BAC
11 clone, 3B-100-L17, is a known 3B centromeric BAC placed in the centromere of 3B
12 chromosome by megabase sequencing analysis, which is highly collinear to the
13 centromere of rice chromosome 1 (Choulet et al. 2010). An extensive comparison of
14 centromeric sequences and distribution of *CR* elements among rice, maize, wheat, and
15 *Brachypodium* might be needed for a complete understanding of the molecular and
16 evolutionary mechanisms underlying the conserved function of centromeres in cereal
17 species.

18

19 Acknowledgments

20

21 We thank Drs Zhao Liu and Gerald Seiler for critical review of the manuscript. This
22 research was supported by a special USDA-NIFA grant to the Wheat Genetic Resources
23 Center, Kansas State University, USA, and fund for excellent young scholar of Shandong

1 Province of China (BS2011SW027).

1 References

- 2
- 3 Abbo S, Dunford RP, Foote T, Reader SM, Flavell RB, Moore G (1995) Organization of
4 retroelement and stem-loop repeat families in the genomes and nuclei of cereals.
5 Chromosome Res 3:5–15
- 6 Ananiev EV, Phillips RL, Rines HW (1998) Chromosome-specific molecular
7 organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci
8 USA 95:13073–13078
- 9 Aragón-Alcaide L, Miller T, Schwatzacher T, Reader S, Moore G (1996) A cereal
10 centromeric sequence. Chromosoma 105:261–268
- 11 Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J (2009)
12 The ‘inner circle’ of the cereal genomes. Curr Opin Plant Bio 12:119–125
- 13 Buchmann JP, Matsumoto T, Stein N, Keller B, Wicker T (2012) Inter-species sequence
14 comparison of *Brachypodium* reveals how transposon activity corrodes genome
15 colinearity. Plant J 71:550–56
- 16 Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu MH, Blattner FR, Jiang J (2002)
17 Functional rice centromeres are marked by a satellite repeat and a centromere-
18 specific retrotransposon. Plant Cell 14:1691–1704
- 19 Choulet F, Wicker T, Rustenholz C et al (2010) Megabase level sequencing reveals
20 contrasted organization and evolution patterns of the wheat gene and transposable
21 element spaces. Plant Cell 22:1686–1701
- 22 Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S et al (1999)
23 Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science
24 286:2468–2474.
- 25 Dong F, Miller T, Jacson SA, Wang GL, Ronald PC, Jinag J (1998) Rice
26 (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci
27 USA 95:8135–8140
- 28 Du J, Z Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, RC Shoemaker RC, Ma J
29 (2010) Evolutionary conservation, diversity and specificity of LTR-
30 retrotransposons in flowering plants: insights from genome-wide analysis and
31 multi-specific comparison. Plant J 63:584–598

- 1 Francki, MG (2001) Identification of *Bilby*, a diverged centromeric Ty1-copia
2 retrotransposon family from cereal rye (*Secale cereale* L.). *Genome* 44:266–274
- 3 Gao D, Gill N, Kim HR, Walling JG, Zhang W, Fan C, Yu Y, Ma J, SanMigue P, Jiang
4 N, et al (2009) A lineage-specific centromere retrotransposon in *Oryza*
5 *brachyantha*. *Plant J* 60:820–831
- 6 Hall SE, Kettler G, Preuss D (2003) Centromere satellites from *Arabidopsis* populations:
7 Maintenance of conserved and variable domains. *Genome Res.* 13:195–205
- 8 Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with
9 rapidly evolving DNA. *Science* 203:1098–1102
- 10 Hudakova S, Michalek W, Presting GG, Ten Hoopen R, Dos Santos K, Jasencakova Z,
11 Schubert I (2001) Sequence organization of barley centromeres. *Nucleic Acids*
12 *Res* 29:5029–5035
- 13 Huo N, Gu YQ, Lazo GR, Vogel JP, Coleman-Derr D, Luo MC, Thilmony R, Garvin DF,
14 Anderson OD (2006) Construction and characterization of two BAC libraries
15 from *Brachypodium distachyon*, a new model for grass genomics. *Genome*
16 49:1099–1108
- 17 Jiang J, Nasuda S, Dong F, Scherrer CW, Woo SS, Wing RA, Gill BS, Ward DC (1996)
18 A conserved repetitive DNA element located in the centromeres of cereal
19 chromosomes. *Proc Natl Acad Sci USA* 93:14210–14213
- 20 Jiang J, Birchler JA, Parrott WA, Dawe RK (2003) A molecular view of plant
21 centromeres. *Trends Plant Sci* 8:570–575
- 22 Jin WW, Lamb JC, Vega JM, Dawe RK, Birchler JA, Jiang J (2005) Molecular and
23 functional dissection of the maize B chromosome centromere. *Plant Cell*
24 17:1412–1423
- 25 Jin WW, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize
26 centromeres: organization and functional adaptation in the genetic background of
27 oat. *Plant Cell* 16:571–581
- 28 Kulikova O, Geurts R, Lamine M et al (2004) Satellite repeats in the functional
29 centromere and pericentromeric heterochromatin of *Medicago truncatula*.
30 *Chromosoma* 113:276–283
- 31 Kumekawa N, Ohtsubo E, Ohtsubo H (1999) Identification and phylogenetic

1 analysis of *gypsy*-type retrotransposons in the plant kingdom. *Genes Genet Syst*
2 74:299–307

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

7 Lamb JC, Yu W, Han F, Birchler JA (2008) Plant centromeres. *Genome Dyn* 4:95–107

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

11 Lim K-B, Yang T-J, Hwang Y-J et al (2007) Characterization of the centromere and peri-
12 centromere retrotransposons in *Brassica rapa* and their distribution in related
13 *Brassica* species. *Plant J* 49:173–183

14 Li B, Choulet F, Heng Y et al (2013) Wheat centromeric retrotransposons: the new ones
15 take a major role in centromeric structure. *Plant J* 73:952–965

16 Liu Z, Yue W, Li D et al (2008) Structure and dynamics of retrotransposons at
17 wheat centromeres and pericentromeres. *Chromosoma* 117:445–456

18 Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes.
19 *Proc Natl Acad Sci USA* 101:12404–12410

20 Ma J, Wing R, Bennetzen JL, Jackson SA (2007) Evolutionary history and positional
21 shift of a rice centromere. *Genetics* 177:1217–1220

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

24 Nagaki K, Song J, Stupar RM et al (2003a) Molecular and cytological analyses of
25 large tracks of centromeric DNA reveal the structure and evolutionary dynamics
26 of maize centromeres. *Genetics* 163:759–770

27 Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J (2003b) Chromatin
28 immunoprecipitation reveals that the 180-bp satellite repeat is the key functional
29 DNA element of *Arabidopsis thaliana* centromeres. *Genetics* 163:1221–1225
30

- 1 Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR,
2 Jiang J (2004) Sequencing of a rice centromere uncovers active genes. *Nature*
3 *Genet* 36:138–145
- 4 Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, Cheng Z, Jiang J (2005)
5 Structure, divergence, and distribution of the CRR centromeric retrotransposon
6 family in rice. *Mol Biol Evol* 22:845–855
- 7 Price AL, Jones NC, Pevzner PA. (2005) De novo identification of repeat families in
8 large genomes. *Bioinformatics* 21 Suppl 1:i351–358
- 9 Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A Ty3-gypsy retrotransposon-like
10 sequence localizes to the centromeric regions of cereal chromosomes. *Plant J*
11 16:721–728
- 12 Qi LL, Friebe B, Zhang P, Gill BS (2009) A molecular-cytogenetic method for locating
13 genes to pericentromeric regions facilitates a genome-wide comparison of synteny
14 between the centromeric regions of wheat and rice. *Genetics* 183:1235–1247
- 15 Qi LL, Friebe B, Wu J, Gu Y, Qian C, Gill BS (2010) The compact *Brachypodium*
16 genome conserves centromeric regions of a common ancestor with wheat and
17 rice. *Funct Integr Genomics* 10:477–492
- 18 Round EK, Flowers SK, Richards EJ (1997) *Arabidopsis thaliana* centromere
19 regions: genetic map positions and repetitive DNA structure. *Genome Res*
20 9:1007–1019
- 21 Šafář J, Bartoš J, Janda J, et al (2004) Dissecting large and complex genomes: flow
22 sorting and BAC cloning of individual chromosomes from bread wheat. *Plant J*
23 39:960–968
- 24
- 25 Sharma A, Presting GG (2008) Centromeric retrotransposon lineages predate the
26 maize/rice divergence and differ in abundance and activity. *Mol Genet Genomics*
27 279:133–147
- 28 Tek AL, Kashihara K, Murata M, Nagaki K (2010) Functional centromeres in soybean
29 include two distinct tandem repeats and a retrotransposon. *Chromosome Res*
30 18:337–347
- 31 The International Brachypodium Initiative (2010) Genome sequencing and analysis of

- 1 the model grass *Brachypodium distachyon*. Nature 463:763–768
- 2 Weber B, Schmidt T (2009) Nested Ty3-gypsy retrotransposons of a single
3 Betaprocumbens centromere contain a putative chromodomain. Chromosome Res
4 17:379-396
- 5 Wen R, Moore G, Shaw PJ (2012) Centromeres cluster de novo at the beginning of
6 meiosis in *Brachypodium distachyon*. PLoS One 7:e44681
- 7 Wu J, Fujisawa M, Tian Z et al (2009) Comparative analysis of complete orthologous
8 centromeres from two subspecies of rice reveals rapid variation of centromere
9 organization and structure. Plant J 60:805–819
- 10 Zhang P, Li WL, Fellers J, Friebe B, Gill BS (2004) BAC-FISH in wheat
11 identifies chromosome landmarks consisting of different types of transposable
12 elements. Chromosoma 112:288–299
- 13 Zhang Y, Huang Y, Zhang L et al (2004) Structural features of the rice chromosome 4
14 centromere. Nucleic Acids Res. 32:2023–2030
- 15 Zhao X, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length
16 LTR retrotransposons. Nucleic Acids Res 35 (Web Server issue): W265–W268
- 17 Zhong CX, Marshall JB, Topp C et al (2002) Centromeric retroelements and satellites
18 interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Table 1 List of BAC clones selected for BAC-FISH

Probed by	BAC clones	Contigs or chromosome	Classification	Reference
Hi10	DH007B23	Ctg118	<i>Brachypodium</i> BAC	This research
Hi10	DH021M4	Ctg271	<i>Brachypodium</i> BAC	This research
Hi10	DH008A23	Ctg290	<i>Brachypodium</i> BAC	This research
Hi10	DH010C12	Ctg42	<i>Brachypodium</i> BAC	This research
Hi10	DH029E4	Ctg42	<i>Brachypodium</i> BAC	This research
Hi10	DH010J10	Singleton	<i>Brachypodium</i> BAC	This research
Hi10	DH010O24	Singleton	<i>Brachypodium</i> BAC	This research
Hi10	DH017I23	Singleton	<i>Brachypodium</i> BAC	This research
pRCS1	DH054L6	Singleton	<i>Brachypodium</i> BAC	This research
pRCS1	DH064P10	Singleton	<i>Brachypodium</i> BAC	This research
pRCS1	DH070K6	Singleton	<i>Brachypodium</i> BAC	This research
pRCS1	DH078K1	Singleton	<i>Brachypodium</i> BAC	This research
pRCS1	DH085J19	Singleton	<i>Brachypodium</i> BAC	This research
pRCS1	DH086J9	Singleton	<i>Brachypodium</i> BAC	This research
BG313557-3L [†]	DH017G05	Bd 2	<i>Brachypodium</i> BAC	Qi et al. 2010
	DH039C01	Bd 2	<i>Brachypodium</i> BAC	Qi et al. 2010
BE637507-4L [†]	DB069J23	Bd 4	<i>Brachypodium</i> BAC	Qi et al. 2010
BE405809-6S [†]	DB042E22	Bd 3	<i>Brachypodium</i> BAC	Qi et al. 2010
BE405195-6S [†]	DB088O14	Bd 3	<i>Brachypodium</i> BAC	Qi et al. 2010
pAet6-09	3B-100-L17	Ctg796	Wheat 3B BAC	Qi unpublished data
	3B-40-L07	Ctg796	Wheat 3B BAC	Qi unpublished data
BF202706-4DL	21E12	NA	<i>Ae. speltoides</i> BAC	Qi et al. 2009
	256K19	NA	<i>Ae. speltoides</i> BAC	Qi et al. 2009
BE497309-4DS	HD008H01	Singleton	<i>Ae. tauschii</i> BAC	Qi et al. 2009

[†] wheat pericentromeric EST. *Brachypodium* BAC clone was selected based on the sequence similarity to the wheat EST.

1

Table 2 The results of BAC-fluorescence *in situ* hybridization (FISH) of *Brachypodium* BAC clones on the mitotic chromosome complements of *Brachypodium*, rye, wheat, barley, rice, and maize.

Probed by	BAC clones	BAC-FISH signal					
		B.d21	Rye	wheat	Barley	Rice	Maize
Hi10	DH021M4	++++	+++	+++	+++	-	-
Hi10	DH029E4	++++	+++	+++	++	-	-
Hi10	DH010C12	++++	+++	++	+	-	-
Hi10	DH007B23	+++	++	+	-	-	-
Hi10	DH008A23	++++	++	+	-	-	-
Hi10	DH017I23	+++	++	NA	NA	NA	NA
pRCS1	DH054L6	+++	++	NA	NA	NA	NA
pRCS1	DH085J19	++++	++	NA	NA	NA	NA
pRCS1	DH086J9	++++	++	NA	NA	NA	NA
Hi10	DH010J10	+++	+	NA	NA	NA	NA
Hi10	DH010O24	+++	+	NA	NA	NA	NA
pRCS1	DH064P10	++++	+	NA	NA	NA	NA
pRCS1	DH070K6	++++	+	NA	NA	NA	NA
pRCS1	DH078K1	++++	+	NA	NA	NA	NA
BG313557-3L [†]	DH039C01	++++	+	NA	NA	NA	NA
BE637507-4L [†]	DB069J23	++++	+	NA	NA	NA	NA
BG313557-3L [†]	DH017G05	++++	-	-	-	-	-
BE405809-6S [†]	DB042E22	++++	-	-	-	-	-
BE405195-6S [†]	DB088O14	++++	-	-	-	-	-

- and + represent, respectively, the absence and presence of hybridization signals: +++++, very strong signal; +++, strong signal; ++, weak signal; +, very weak signal.

[†]wheat pericentromeric EST. *Brachypodium* BAC clone was selected based on the sequence similarity to the wheat EST.

2 NA: not apply.

3

1

2 Table 3 The results of BAC-fluorescence *in situ* hybridization (FISH) of Triticeae BAC clones on
 3 the mitotic chromosome complement of wheat Chinese Spring (CS) and *B. distachyon* (B.d21)

BAC clones	BAC-FISH signal	
	CS	B.d21
3B-40-L07	+++	++
3B-100-L17	+++	-
21E12	+++	-
256K19	+++	-
HD008H01	+++	-

4 - and + represent, respectively, the absence and presence of hybridization signals: +++,
 5 strong signal; ++, weak signal.

6

7 Table 4 Common LTR retrotransposons identified in the four *Brachypodium* BACs

No.	Name	Family	Structure	Total length of conserved sequence (kb)				
				DB069J23	DB088O14	DB042E22	DH017G05	whole genome
1	<i>CRBd1</i> (9.76 kb)	<i>Gypsy</i>	truncated	101.2	44.6	7.5	20.8	4,410.6 [†]
2	<i>CRBd2</i> (7.23 kb)	<i>Gypsy</i>	Full length, similar to rice <i>CRR3</i>	7.6 (Full length)	7.2 (Full length)	0.3	0.9 (solo LTR)	263.7
3	<i>REBd1</i> (11.64kb)	<i>Gypsy</i>	Full length	2.2	22.2	11.3	0.9 (solo LTR)	2,011.1
4	<i>REBd2</i> (7.29 kb)	<i>Gypsy</i>	Internal coding region	15.1	0	12.4	11.9	4264.0
5	<i>REBd3</i> (8.03 kb)	<i>Copia</i>	Full length	0	0	8.0 (Full length)	8.0 (Full length)	913.5

8
9

[†] Of them, the total length of sequences conserved to the LTR region of *CRBd1* was 2758.7 kb.

1
2
3

Table S1 Hybridization results of positive *Brachypodium* BACs with centromeric-specific clones, Hi10, pAet6-09, and pRCS1

Probed by	Associated BAC	No. fragments of BAC hybridizing to:			
		Hi10	pAet6-09	pRCS1	
Hi10	DH002F21	8	8	2	
	DH007B23 [†]	2	2	0	
	DH007C24	7	7	1	
	DH008A23 [†]	8	8	3	
	DH010C12 [†]	6	7	1	
	DH010J10 [†]	10	10	1	
	DH010O24 [†]	14	16	3	
	DH026L22	4	2	2	
	DH014I5	2	4	2	
	DH017I23 [†]	7	7	2	
	DH021M4 [†]	2	3	1	
	DH024H22	9	11	1	
	DH029E4 [†]	5	5	1	
	pRCS1	DH003K1	3	5	1
		DH011N9	5	6	2
		DH014I7	7	9	2
DH017C21		7	6	4	
DH027C19		0	1	2	
DH028M13		9	9	3	
DH030F13		4	2	2	
DH031H5		3	3	3	
DH032L2		3	3	2	
DH032J23		1	1	2	
DH039M8		9	9	3	
DH042I3		5	5	3	
DH054L6 [†]		10	9	3	
DH056C10		5	5	2	
DH060A1		5	8	1	
DH062J15		6	7	2	
DH064P10 [†]		11	12	2	
DH070K6 [†]		10	10	2	
DH078K1 [†]		12	12	4	
DH085J19 [†]		15	13	2	
DH086J9 [†]	14	13	4		
DH087M5	3	4	1		
DH089N24	9	9	1		
DH090B7	5	6	1		
DH090F5	8	8	3		

[†] selected for BAC-FISH

4
5

1
2

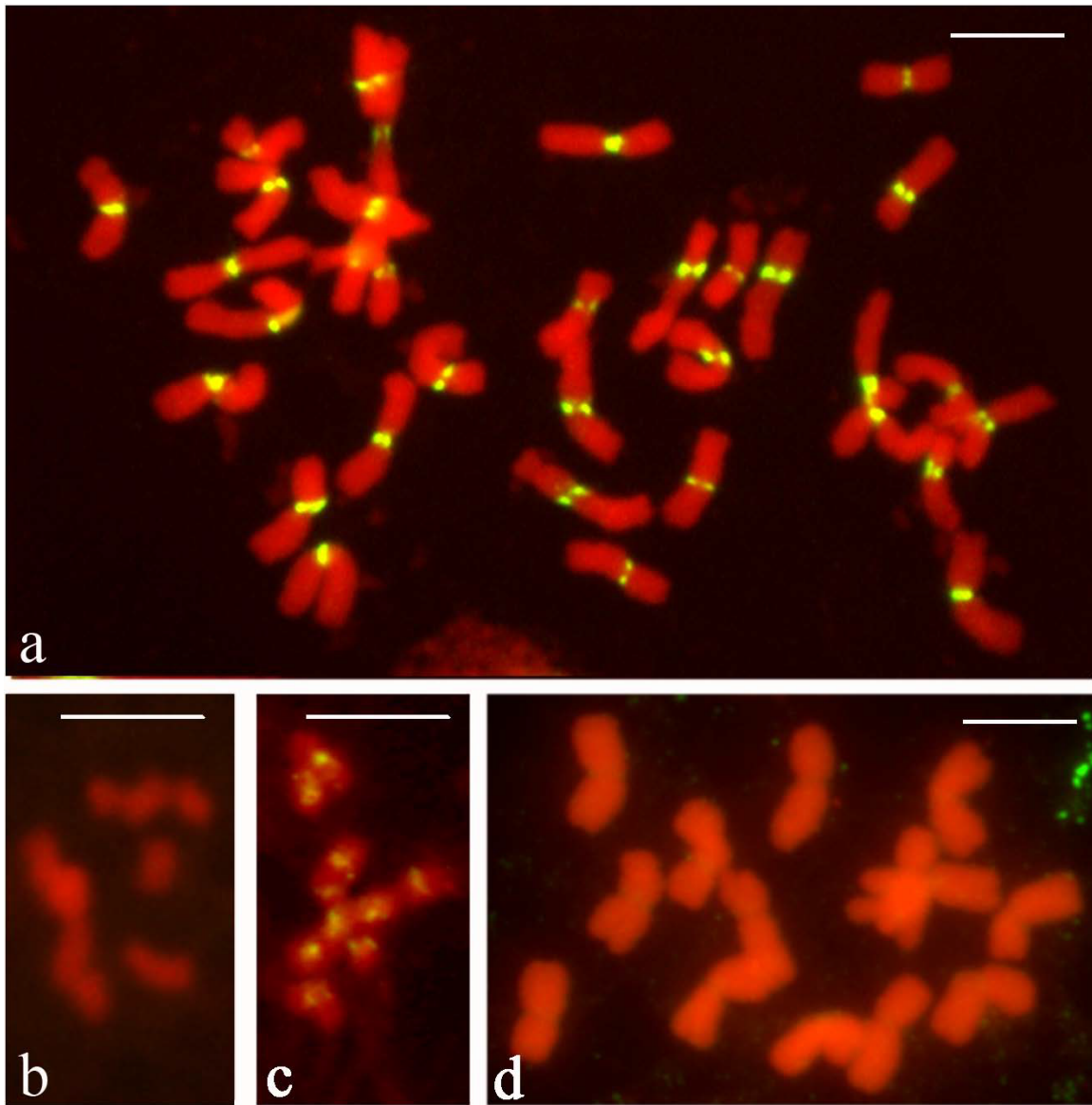
Table S2 Distribution of the full-length *CRBdI* in *B. distachyon* genome

Chromosome No.	Position of Bd_CENT repeats in centromere regions	Position of Bd_CNET repeats in chromosome fusion points	Position of Bd_CNET repeats outside of the centromeres	Position of full length <i>CRBdI</i> (Insertion time)
1	37.379-38.177 Mb (1294) [†]	24.696-24.700 Mb (27)	23.017-23.020 Mb (13),	39.080-39.093 Mb (0.14Myr)
		50.742-50.744 Mb (13)	30.402 Mb (2)	70.916-70.929 Mb (0.01Myr)
			36.856-36.866 Mb (47)	
			40.922 Mb (3)	
2	28.989-29.716 Mb (340)	12.733-12.735 Mb (13)		26.344-26.357 Mb (0.06Myr)
		40.087-40.088 Mb (12)		
3	25.158-25.675 Mb (1607)	11.136-11.153 Mb (28)	24.365-24.388 Mb (53)	
4	20.641-21.023 Mb (1264)	24.724-24.734 Mb (61)		8.206-8.219 Mb (0.02Myr)
				22.075-22.088 Mb (1.34Myr)
				26.875-26.888 Mb (0.01Myr)
5	7.608-7.731 Mb (194)		7.293-7.314 Mb (14)	1.524-1.537 Mb (0.12Myr)
			8.103 Mb (1)	9.032-9.045 Mb (0.14Myr)

[†]The numbers in parentheses represent the copy numbers of Bd_CENT

3

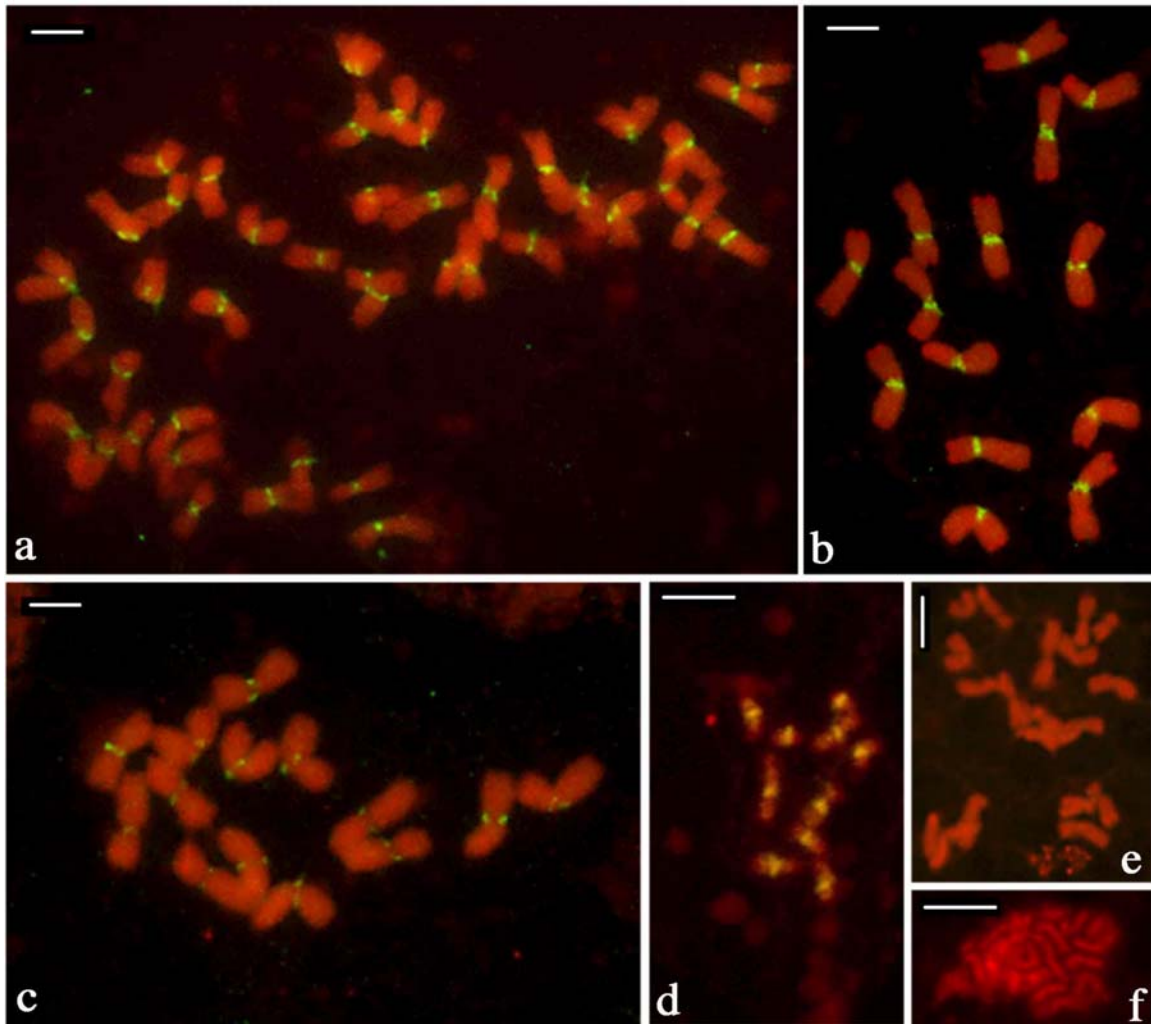
- 1 Figure legends
- 2 Figure 1



- 3
- 4 Fig. 1 Wheat 3B BAC clone 3B-100-L17 hybridized to mitotic chromosomes of wheat
- 5 (a) and *Brachypodium* (b). No FISH signal was observed in *Brachypodium* chromosomes
- 6 (b). *Brachypodium* BAC clone DB088O14 hybridized to mitotic chromosomes of
- 7 *Brachypodium* (c) and rye (d). No FISH signal was observed in rye chromosomes (d), as
- 8 well as in wheat, barley, maize, and rice (data not shown). Scale bar is 5 μm .

9

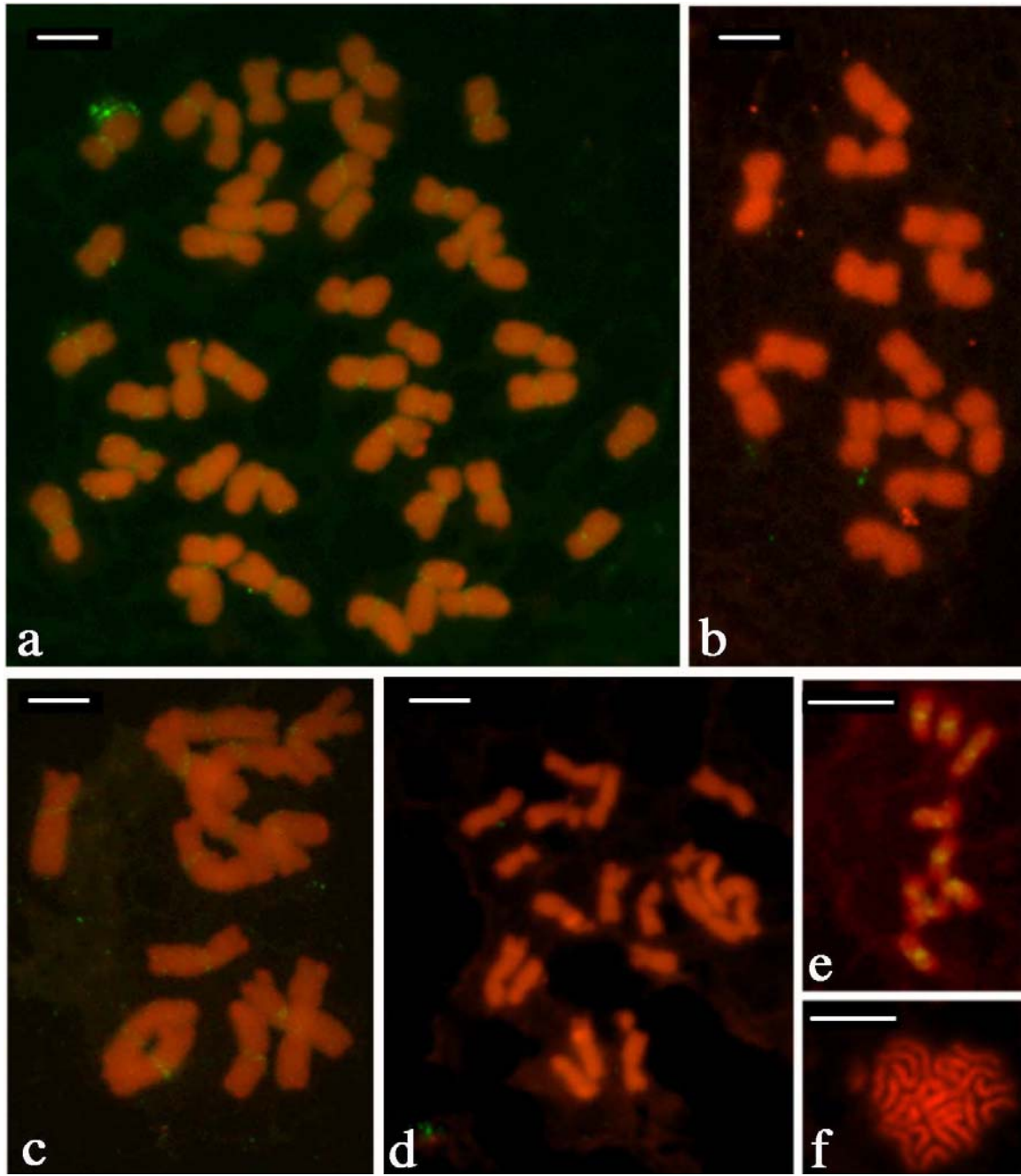
1 Figure 2



2
3 Fig. 2 *Brachypodium* BAC clone DH029E4 hybridized to mitotic chromosomes of wheat
4 (a), rye (b), barley (c), *Brachypodium* (d), Maize (e), and rice (f). No FISH signal was
5 observed in maize (e) and rice (f) chromosomes. Scale bar is 5 μ m.

6

1 Figure 3

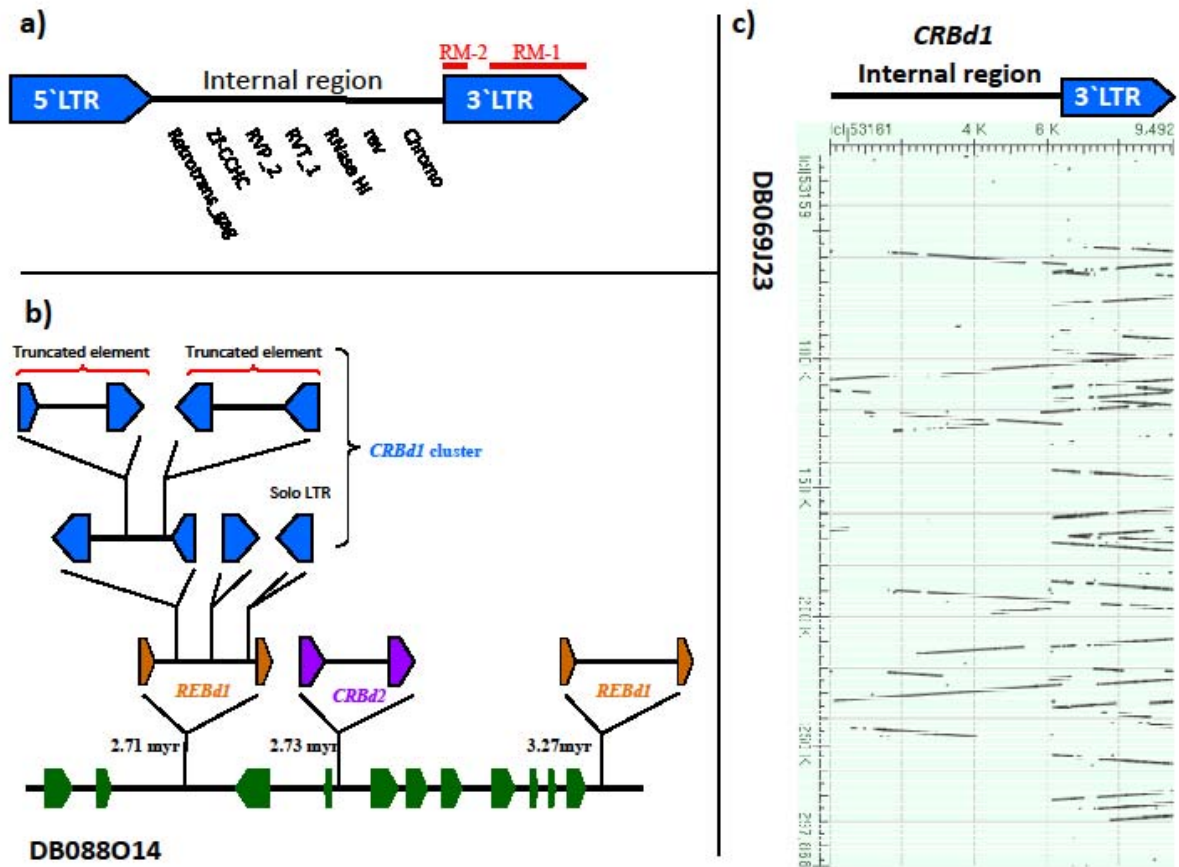


2

3 Fig. 3 *Brachypodium* BAC clone DH007B23 hybridized to mitotic chromosomes of
4 wheat (a), barley (b), rye (c), maize (d), Bd21 (e), and rice (f). No FISH signal was
5 observed in the chromosomes of barley (b), maize (d), and rice (f). Scale bar is 5 μ m.

6

1 Figure 4



2
3

4 Fig. 4 The structure and distribution of the retrotransposon *CRBd1*. a) The full length
5 element of *CRBd1* retrotransposon in *Brachypodium* was 12.5-12.8 kb in length with
6 LTRs of 3.1-3.3kb. The deduced coding sequence contains typical domains of
7 retrotransposon: gag protein (Retrotrans_gag), pol protein (RVP_2), reverse transcriptase
8 (RVT_1), RNase H1 and integrase (rev), as well as other two domains, the Zinc knuckle
9 domain (zf-CCHC) and Chromatin Organization Modifier domain (Chromo). The
10 *Brachypodium* specific repeats, RM-1 and RM-2 identified in DB069J23, were fragments
11 of the LTR region of *CRBd1*. b) Distribution of the retrotransposons in the BAC

1 DB088O14. The green boxes represent the 11 gene models in the 200kb-region of
2 chromosome 3:46336326..46536325. Black lines flanked by two boxes indicate
3 retrotransposons, under which the names were labeled. Retrotransposons were inserted
4 into other retrotransposons or intergenic region. A *CRBdl* cluster was noticed in this
5 BAC, including the truncated elements and solo-LTRs. The insertion time of full length
6 retrotransposons were calculated and labeled under the elements. c) The dot matrix view
7 of the alignment between BAC DB069J23 (Bd4: 22540676..22840675) and the
8 retrotransposon *CRBdl*(DB088O14:25723..35498), indicating the distribution of the
9 sequences with similarity to *CRBdl* in this BAC. For simplicity, only the internal region
10 and the 3'LTR of *CRBdl* were used for comparison.

11

1 Figure 5

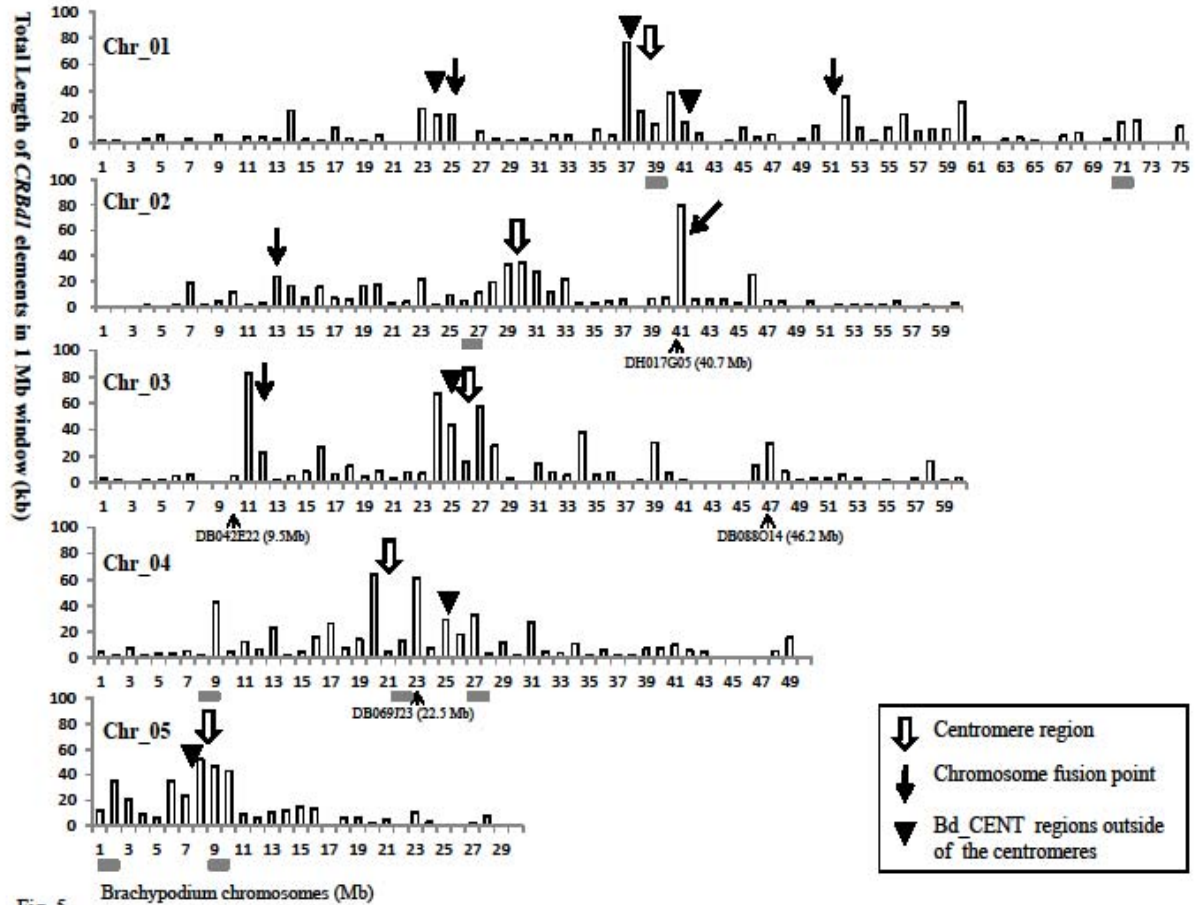


Fig. 5

2

3

4 Fig. 5 The distribution histograms of the LTR of *CRBd1* on the *Brachypodium*

5 chromosomes. The total sequence length (kb) of alignments in 1Mb window was plotted

6 along the chromosomes. The positions of centromeres, chromosome fusion points, and

7 Bd_CENT containing regions outside of the centromeres were marked. The arrows point

8 the positions of the four *Brachypodium* BACs, and the gray bars represent the positions

9 of the eight full-length *CRBd1*.

10

1 Figure 6

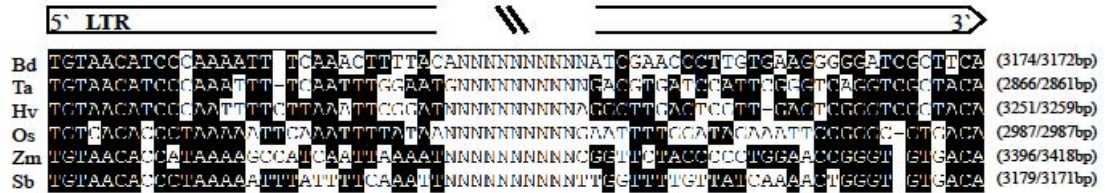


Fig. 6

2
3

4 Fig. 6 Sequence alignment of the LTRs from different species. Full length
 5 retrotransposons were identified in species of *Brachypodium* (Bd, chr1_
 6 70915934..70928501), wheat (Ta, FN564426_562537..573702), barley (Hv,
 7 AC250228_46960..58979), rice (Os, AP008246_54271..66260), maize (Zm,
 8 AF448416_48249..61044) and sorghum (Sb, chr9_9757379..9770073) respectively. The
 9 boundaries (30 bp from the 5' and 3' end, respectively) of the LTRs were used for
 10 alignment. The numbers on right indicate the length of 5' and 3' LTR for each
 11 retrotransposon.

12

1 Figure S1

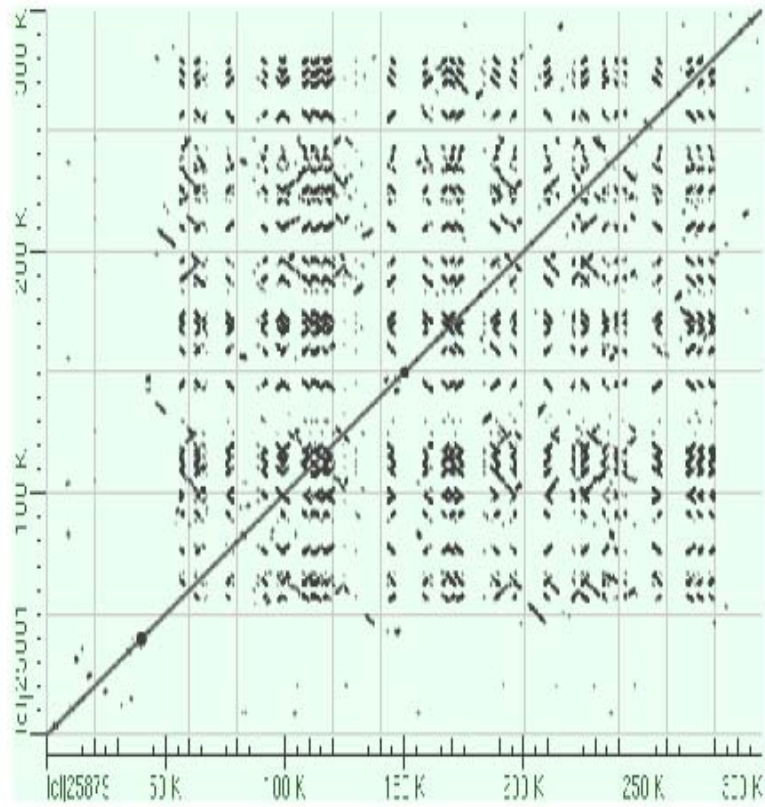


Fig. S1

2
3 Fig. S1 The dot matrix view of self alignment of the BAC DB069J23.