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

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.

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

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

L.L. Qi[1*†], J.J Wu[2*], B. Friebe[3], C. Qian[3], Y.Q. Gu[4], D.L. Fu[2], and B.S. Gill[3]

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

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

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

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

*These authors contributed equally to this work

[†]For correspondence (phone +1 701 239 1351; fax +1 701 239 1346; email lili.qi@ars.usda.gov)

Abstract

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

Introduction

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

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

Unlike centromeric satellites, the *CR* family in grass species is highly conserved. Two highly conserved *CR* sequences, *CCS1* and *pSau3A9,* which are parts of Ty3/*gypsy*-type retrotransposons, were first found to localize at the centromeres of most cereal species that have been investigated (Aragon-Alcaide et al.1996; Jiang et al 1996, Miller et al. 1998; Presting et al. 1998). *CRR* (*CR* of rice) and *CRM* (*CR* of maize) are the most intensively studied *CR* elements among plant species (Dong et al. 1998; Presting et al. 1998; Cheng et al. 2002; Zhong et al. 2002; Nagaki et al. 2003a, 2005). Rice *CRR1* is homologous to maize *CRM3*, *CRR2* to *CRM2*, *CRR3* to *CRM1*, and *CRR4* to *CRM4*, which pre-date the divergence of maize and rice (Sharma and Presting 2008). Two putative *CR* families of soybean were also grouped to *CRR* and *CRM* lineage (Du et al. 2010) and the *CR* elements, *Beetle1*and *Beetle 2*, found in beet are highly similar to the *CRs* of rice, maize, and Barley (Weber and Schmidt 2009). The *CR* elements isolated from barley and wheat showed cross hybridization among cereal species (Hudakova et al. 2001; Zhang P et al. 2004). As few exceptions to the general *CR* conservation of grasses, species-specific *CR* element was reported in rye (Francki 2001) and wild rice (Gao et al. 2009). A rye-specific *CR*, *Bilby* that is a Ty1-*copia* retrotransposon-like element, is highly divergent from other known cereal *CR* elements, and a lineage-specific *CR* element was identified in *Oryza brachyantha*.

Brachypodium distachyon (hereafter referred to as *Brachypodium*) is a wild annual grass belonging to the Pooideae, more closely related to wheat, barley, and forage grasses than rice and maize. As an experimental model, the completed genome sequence of *Brachypodium* provides an important reference for grass biology and centromere studies

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

Materials and Methods

Materials

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

Methods

Selection of *Brachypodium* putative centromeric BAC

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

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

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

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

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

Mitotic chromosome spreads for BAC-FISH were prepared from the root tips of *Brachypodium* Bd21, CS wheat, Imperial rye, Betzes barley, Nipponbare rice, and B73 maize as described by Qi et al. (2010). BAC DNA was isolated using a Qiagen Plasmid Midi Kit following the manufacture's instruction (Qiagen Valencia. Calif.). One microgram of BAC DNA was labeled with fluorescein-12-dUTP (Enzo Life Science Inc, Farmingdale, NY) using nick translation. The BAC-FISH was performed on metaphase chromosomes as previously described (Qi et al. 2010). 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).

BAC sequence annotation

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

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

Results

Identification of *Brachypodium*–specific centromeric BACs

The *B. distachyon* BAC library (~4.5 × coverage) was probed with Hi10 and pRCS1clones. Of the 38 unambiguous positive BAC clones, 13 were selected by probe Hi10 and 25 by pRCS1 (Table S1). The BAC clones were digested with restriction enzyme *Hind*III and hybridized to the three centromeric DNA sequences, Hi10, pRCS1, and pAet6-09. Positive Southern hybridization signals and ladder patterns were detected. Average numbers of BAC fragments hybridizing to three clones were 6.7 (range from 0 to 15) for Hi10, 7.0 (1 to 16) for pAet6-09, and 2.0 (0 to 4) for pRCS1 (Table S1). Fourteen BAC clones that gave intense hybridization signals were selected for further BAC-FISH with chromosome complements of *Brachypodium* and rye (Table S1).

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

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

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

BAC-FISH of Triticeae centromeric BAC clones to *Brachypodium* chromosomes

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

Sequence organization of *Brachypodium*–specific centromeric BAC clones

Common repeats in the four Brachypodium BACs

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

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

Detailed analysis of the *Brachypodium* genomic regions containing RM-1 and RM-2 revealed that these two repeats were parts of the LTR of one *Gypsy* retrotransposon found in BAC DB088O14, designated as *CRBd1* (Centromeric Retrotransposon of *B. distachyon*, GeneBank # KF040483) (Fig. 4a). RM-1 and RM-2 were located immediately at the 3' and 5' end of the LTR, respectively. The complete LTR was about 3.3 kb in length, and between the RM-1 and RM-2 was a region with high GC content (~70%). Eight full length copies of *CRBd1* were identified in the *Brachypodium* genome, with size ranging from 12. 5 kb to 12.8 kb (Supplementary file 1), and five of them were located less than 3 Mb from the centromeres or chromosome fusion points (Fig. 5, Table S2). The insertion time of the eight full length *CRBd1* was estimated to be in the range of 0.01 to 1.34 million years ago (Mya). Target site duplications (TSD) were found in 7 of them. In the internal region (~6.2 kb) between two LTRs, there were one Zinc knuckle domain (pfam: zf-CCHC) and one chromatin organization modifier domain (pfam: Chromo), besides feature proteins present in a typical LTR retrotransposable element such as retrotransposase (Fig. 4a). Using the chromodomain identified in *CRBd1*as query sequence, a total of 420 copies (e-value < 1e-5) were identified in whole *Brachypodium* genome. The distribution patterns of chromodomain are well consistent with that of *CRBd1* which is enriched in centromere regions (Fig. 5).

No full length element of *CRBd1*was identified in the four *Brachypodium* BACs. In BAC DB088O14, two truncated and one fragmented elements, and one partial LTR and one solo LTR were clustered in a 65.7 kb region in a different retrotransposon (Fig. 4b). The *CRBd1* homologous sequences occupied a total of 44.6 kb (67.9%) of this region. In

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

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

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

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

Other retrotransposons in the four Brachypodium BACs

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

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

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

Discussion

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

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

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

yielded FISH signals at their original positions. These results imply that the accumulation of the *CR* elements originally present in these BACs have preferentially occurred in the regions of the currently active centromeres. It is also evident that the *CR* element of *CRBd1* is more abundant in BAC DB069J23, which is located at the active centromere region of Bd chromosome 4 (Fig. 5). The conserved sequence of *CRBd1* were 101.2 kb in length in DB069J23 compared to 44.6 kb in DB088O14, 20.8 kb in DH017G05, and 7.5 kb in DB042E22 (Table 4). These results support the hypothesis that redundant centromeres in *Brachypodium* chromosomes became inactive by the loss of centromeric retrotransposons and rapid turnover of centromere-specific satellites (Qi et al. 2010). In other words, the *Brachypodium* active centromeres maintain centromere satellite repeats and accumulate *CR* elements as a result of centromere drive (Ma et al. 2007; Wu et al. 2009). Only eight full length retrotransposons of *CRBd1* were found in the *Brachypodium* genome, and many solo-LTR of *CRBd1* are present in the four BACs analyzed and in the *Brachypodium* genome, revealing that *CRBd1* is an ancient centromeric retrotransposon.

Rice and *Brachypodium* diverged approximately 40-54 MYA, while *Brachypodium* and wheat diverged approximately 30 MYA (The International Brachypodium Initiative 2010). Although their genomes vary in size and basic chromosome numbers, gene content and gene order has been largely conserved. The conserved genes were also reported to be present in the centromere regions of rice, wheat, and *Brachypodium,* which share the syntenic blocks among several sets of homologous centromeres (Qi et al. 2010). However, the *Brachypodium CR* elements appear to be highly divergent from other grass species, especially from rice and maize. Except three BAC clones mentioned above, five

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

Acknowledgments

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

1 Province of China (BS2011SW027).

References

Abbo S, Dunford RP, Foote T, Reader SM, Flavell RB, Moore G (1995) Organization of retroelement and stem-loop repeat families in the genomes and nuclei of cereals. Chromosome Res 3:5–15

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

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 Bio 12:119–125

Buchmann JP, Matsumoto T, Stein N, Keller B, Wicker T (2012) Inter-species sequence comparison of *Brachypodium* reveals how transposon activity corrodes genome colinearity. Plant J 71:550–56

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

Choulet F, Wicker T, Rustenholz C et al (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S et al **(**1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science 286:2468–2474.

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

Du J, Z Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, RC Shoemaker RC, Ma J (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. Plant J 63:584–598

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

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

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

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

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

Jin WW, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. Plant Cell 16:571–581

Kulikova O, Geurts R, Lamine M et al (2004) Satellite repeats in the functional centromere and pericentromeric heterochromatin of Medicago truncatula. Chromosoma 113:276–283

Kumekawa N, Ohtsubo E, Ohtsubo H (1999) Identification and phylogenetic

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

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

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

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

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

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

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

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

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

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

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

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. Nature 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

Price AL, Jones NC, Pevzner PA. (2005) De novo identification of repeat families in large genomes. Bioinformatics 21 Suppl 1:i351–358

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, 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

Qi LL, Friebe B, Wu J, Gu Y, Qian C, Gill BS (2010) The compact *Brachypodium genome* conserves centromeric regions of a common ancestor with wheat and rice. Funct Integr Genomics 10:477–492

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

Šafář J, Bartoš J, Janda J, et al (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J 39:960–968

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

Tek AL, Kashihara K, Murata M, Nagaki K (2010) Functional centromeres in soybean include two distinct tandem repeats and a retrotransposon. Chromosome Res 18:337–347

The International Brachypodium Initiative (2010) Genome sequencing and analysis of

the model grass *Brachypodium distachyon*. Nuture 463:763–768

Weber B, Schmidt T (2009) Nested Ty3-gypsy retrotransposons of a single Betaprocumbens centromere contain a putative chromodomain. Chromsome Res 17:379-396

Wen R, Moore G, Shaw PJ (2012) Centromeres cluster de novo at the beginning of meiosis in *Brachypodium distachyon.* PLoS One 7:e44681

Wu J, Fujisawa M, Tian Z 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

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

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

Zhao X, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35 (Web Server issue): W265–W268

Zhong CX, Marshall JB, Topp C et al (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

1

Table 1 List of BAC clones selected for BAC-FISH

Probed by	BAC clones	Contigs or chromosome	Classification	Reference
Hi10	DH007B23	Ctg118	*Brachypodium* BAC	This research
Hi10	DH021M4	Ctg271	*Brachypodium* BAC	This research
Hi10	DH008A23	Ctg290	*Brachypodium* BAC	This research
Hi10	DH010C12	Ctg42	*Brachypodium* BAC	This research
Hi10	DH029E4	Ctg42	*Brachypodium* BAC	This research
Hi10	DH010J10	Singleton	*Brachypodium* BAC	This research
Hi10	DH010O24	Singleton	*Brachypodium* BAC	This research
Hi10	DH017I23	Singleton	*Brachypodium* BAC	This research
pRCS1	DH054L6	Singleton	*Brachypodium* BAC	This research
pRCS1	DH064P10	Singleton	*Brachypodium* BAC	This research
pRCS1	DH070K6	Singleton	*Brachypodium* BAC	This research
pRCS1	DH078K1	Singleton	*Brachypodium* BAC	This research
pRCS1	DH085J19	Singleton	*Brachypodium* BAC	This research
pRCS1	DH086J9	Singleton	*Brachypodium* BAC	This research
BG313557-3L†	DH017G05	Bd 2	*Brachypodium* BAC	Qi et al. 2010
	DH039C01	Bd 2	*Brachypodium* BAC	Qi et al. 2010
BE637507-4L†	DB069J23	Bd 4	*Brachypodium* BAC	Qi et al. 2010
BE405809-6S†	DB042E22	Bd 3	*Brachypodium* BAC	Qi et al. 2010
BE405195-6S†	DB088O14	Bd 3	*Brachypodium* 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	*Ae. speltoides* BAC	Qi et al. 2009
	256K19	NA	*Ae. speltoides* BAC	Qi et al. 2009
BE497309-4DS	HD008H01	Singleton	*Ae. tauschii* BAC	Qi et al. 2009

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

2
3

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

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

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

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

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	*CRBd1* (9.76 kb)	*Gypsy*	truncated	101.2	44.6	7.5	20.8	4,410.6[†]
2	*CRBd2* (7.23 kb)	*Gypsy*	Full length, similar to rice *CRR3*	7.6 (Full length)	7.2 (Full length)	0.3	0.9 (solo LTR)	263.7
3	*REBd1* (11.64kb)	*Gypsy*	Full length	2.2	22.2	11.3	0.9 (solo LTR)	2,011.1
4	*REBd2* (7.29 kb)	*Gypsy*	Internal coding region	15.1	0	12.4	11.9	4264.0
5	*REBd3* (8.03 kb)	*Copia*	Full length	0	0	8.0 (Full length)	8.0 (Full length)	913.5

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

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

		No. fragments of BAC hybridizing to:		
Probed by	Associated BAC	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

Table S2 Distribution of the full-length *CRBd1* 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 *CRBd1* (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

Figure legends

Figure 1

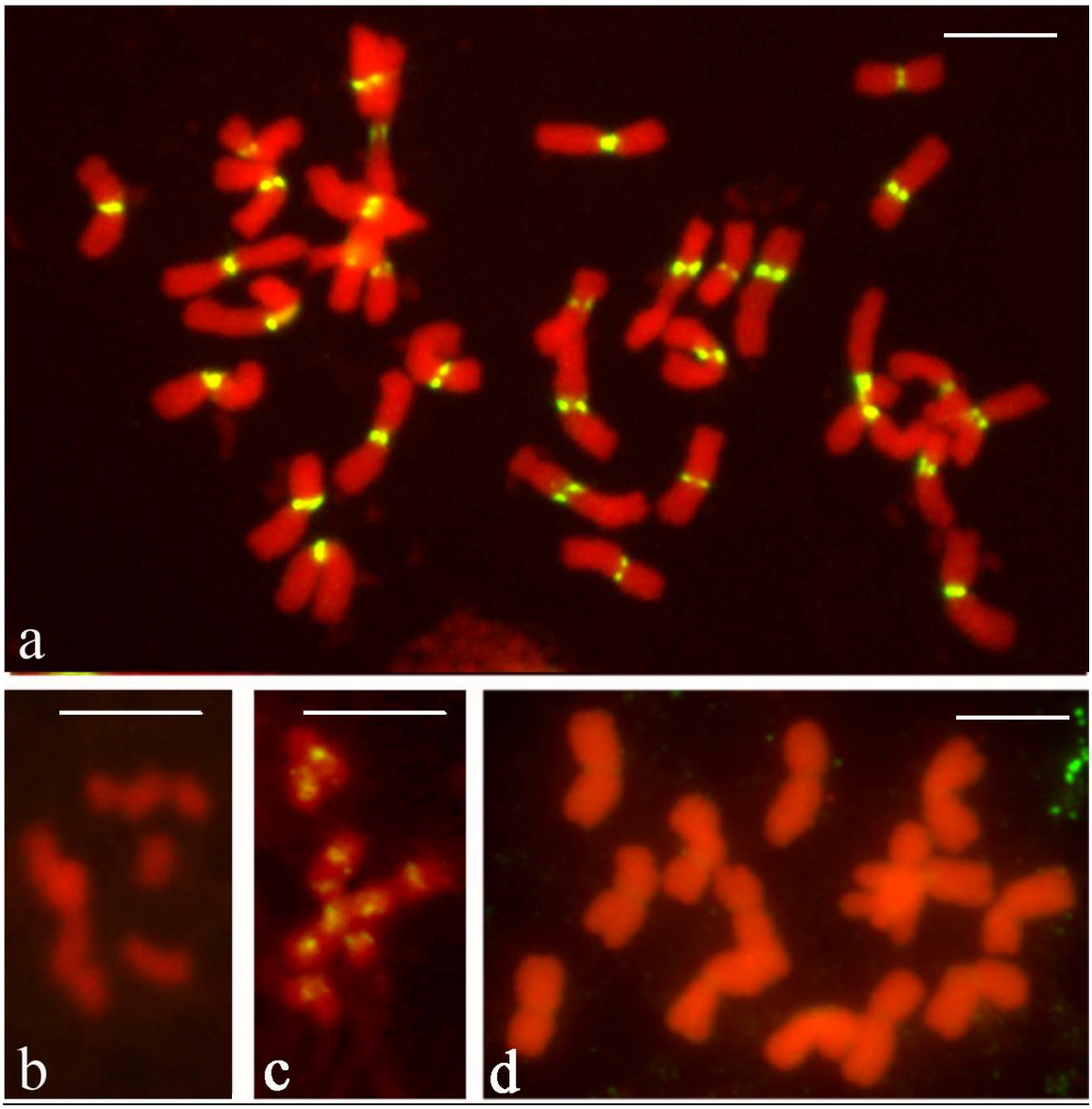

Fig. 1 Wheat 3B BAC clone 3B-100-L17 hybridized to mitotic chromosomes of wheat (a) and *Brachypodium* (b). No FISH signal was observed in *Brachypodium* chromosomes (b). *Brachypodium* BAC clone DB088O14 hybridized to mitotic chromosomes of *Brachypodium* (c) and rye (d). No FISH signal was observed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

Figure 2

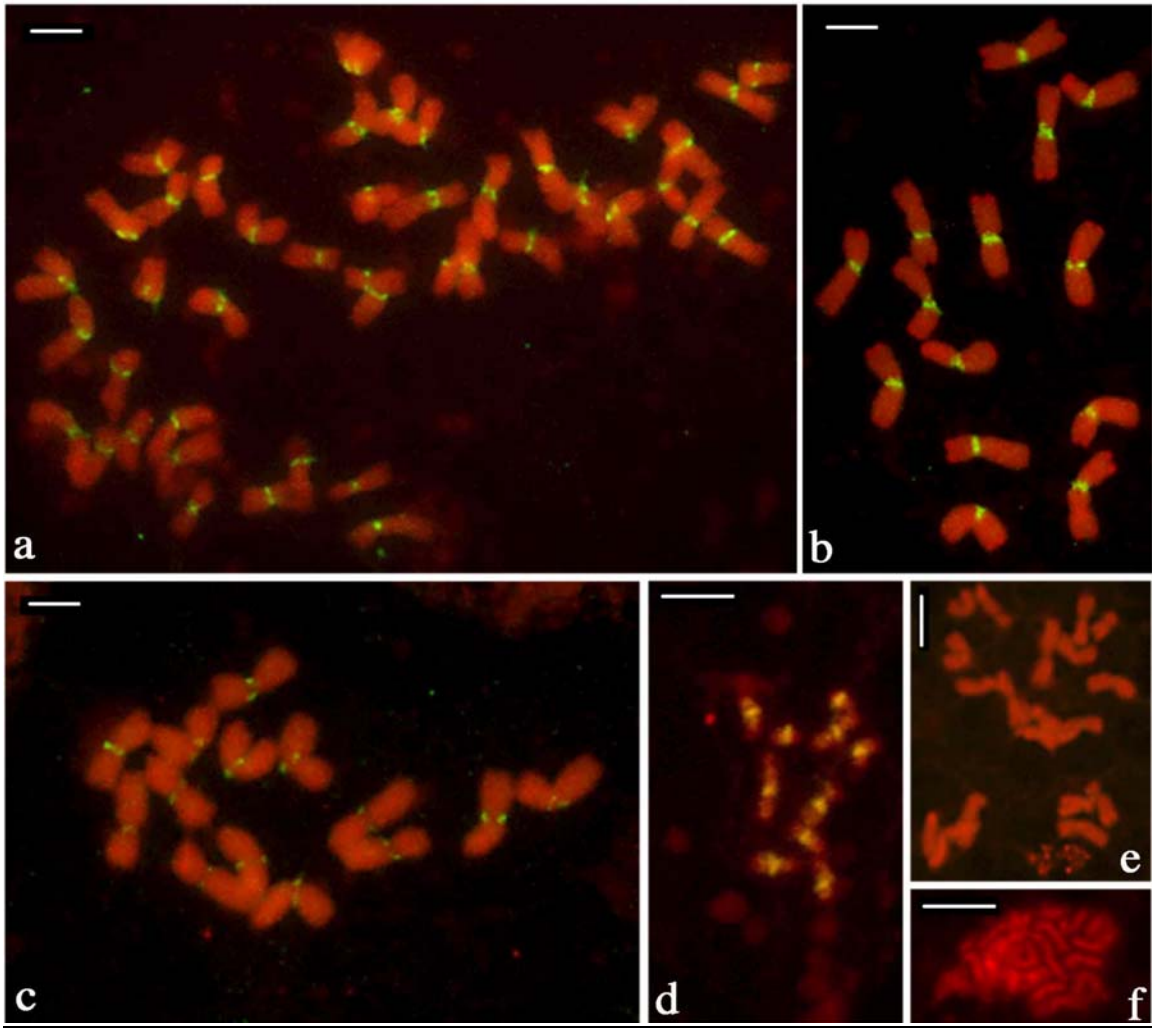

Fig. 2 *Brachypodium* BAC clone DH029E4 hybridized to mitotic chromosomes of wheat (a), rye (b), barley (c), *Brachypodium* (d), Maize (e), and rice (f). No FISH signal was observed in maize (e) and rice (f) chromosomes. Scale bar is 5 μm.

Figure 3

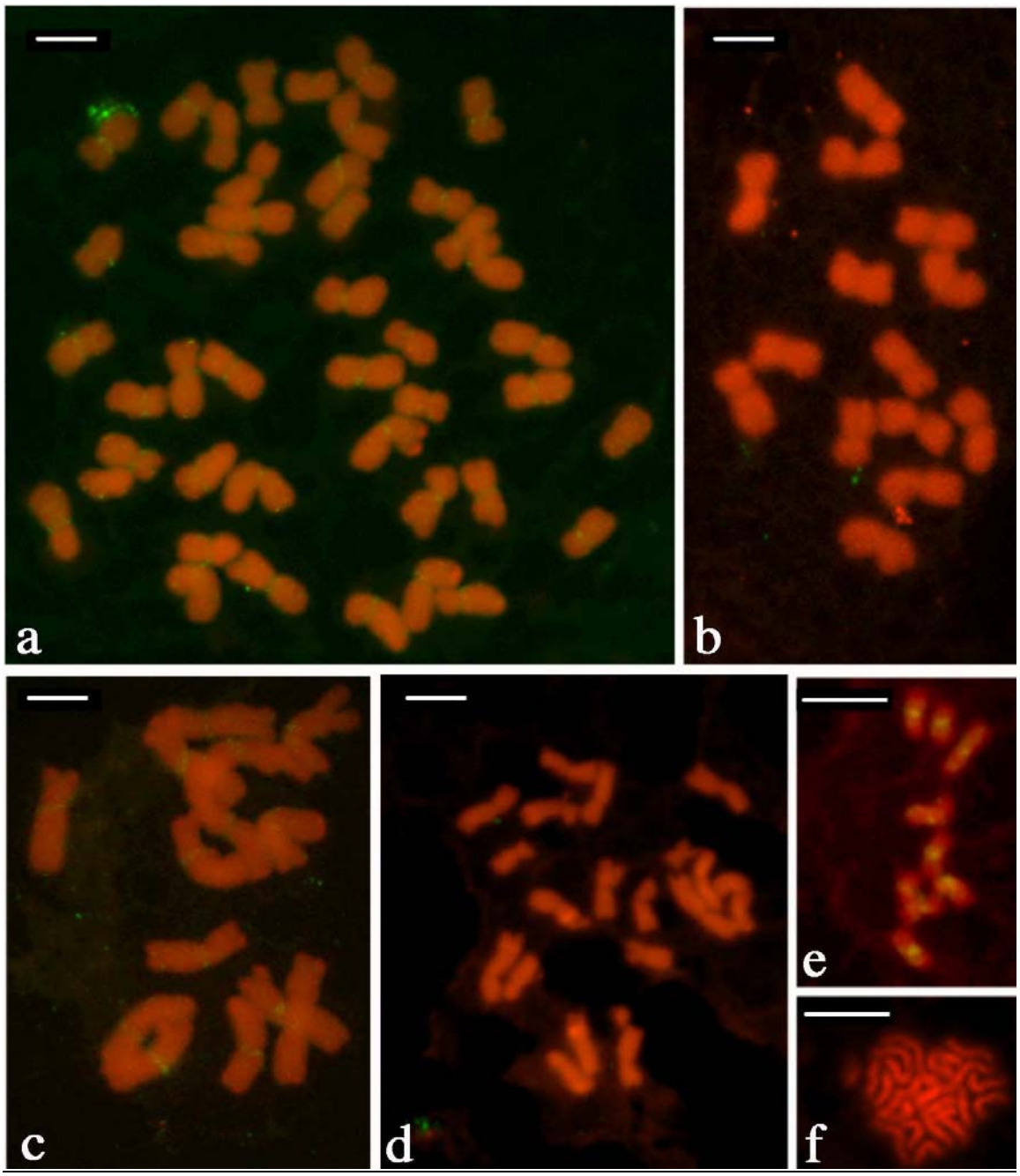

Fig. 3 *Brachypodium* BAC clone DH007B23 hybridized to mitotic chromosomes of wheat (a), barley (b), rye (c), maize (d), Bd21 (e), and rice (f). No FISH signal was observed in the chromosomes of barley (b), maize (d), and rice (f). Scale bar is 5 μm.

Figure 4

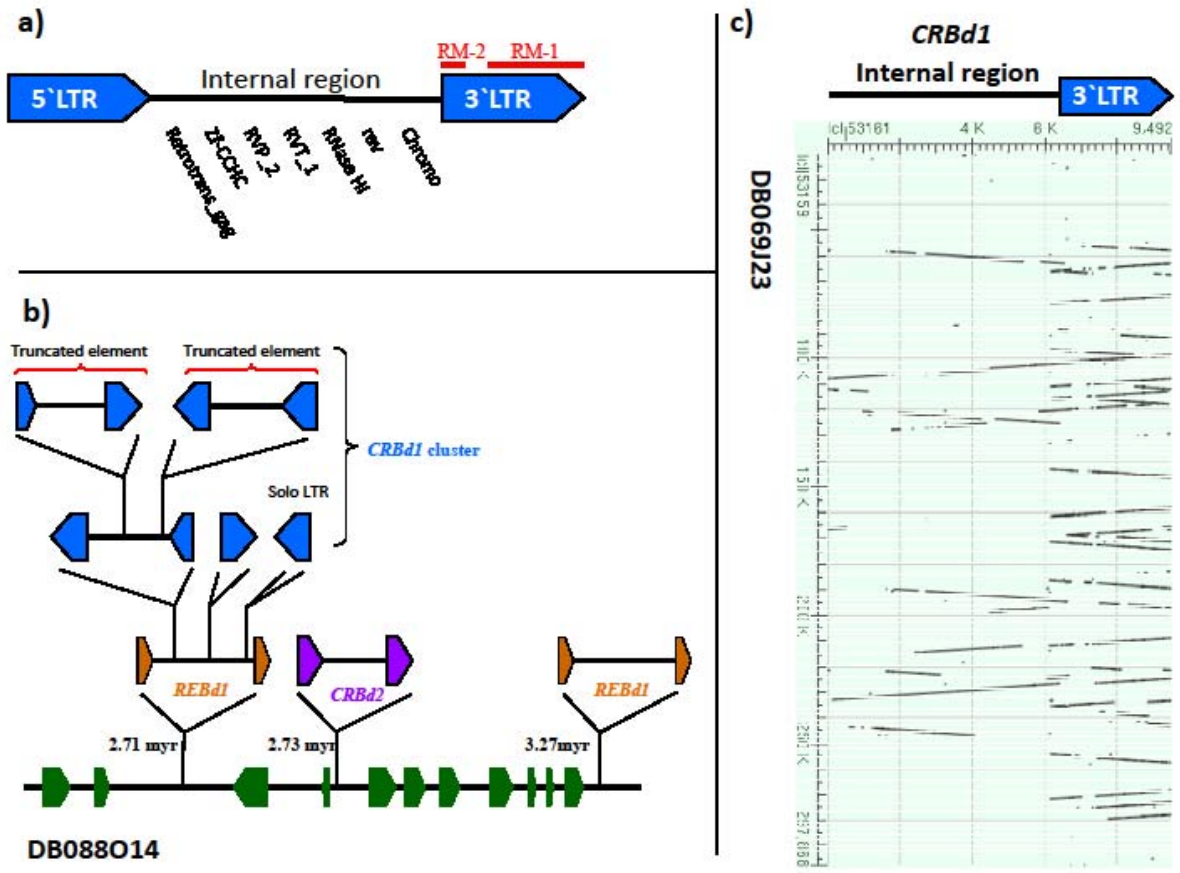

Fig. 4 The structure and distribution of the retrotransposon *CRBd1*. a) The full length element of *CRBd1* retrotransposon in *Brachypodium* was 12.5-12.8 kb in length with LTRs of 3.1-3.3kb. The deduced coding sequence contains typical domains of retrotransposon: gag protein (Retrotrans_gag), pol protein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

DB088O14. The green boxes represent the 11 gene models in the 200kb-region of chromosome 3:46336326..46536325. Black lines flanked by two boxes indicate retrotransposons, under which the names were labeled. Retrotransposons were inserted into other retrotransposons or intergenic region. A *CRBd1* cluster was noticed in this BAC, including the truncated elements and solo-LTRs. The insertion time of full length retrotransposons were calculated and labeled under the elements. c) The dot matrix view of the alignment between BAC DB069J23 (Bd4: 22540676..22840675) and the retrotransposon *CRBd1*(DB088O14:25723..35498), indicating the distribution of the sequences with similarity to *CRBd1* in this BAC. For simplicity, only the internal region and the 3'LTR of *CRBd1* were used for comparison.

Figure 5

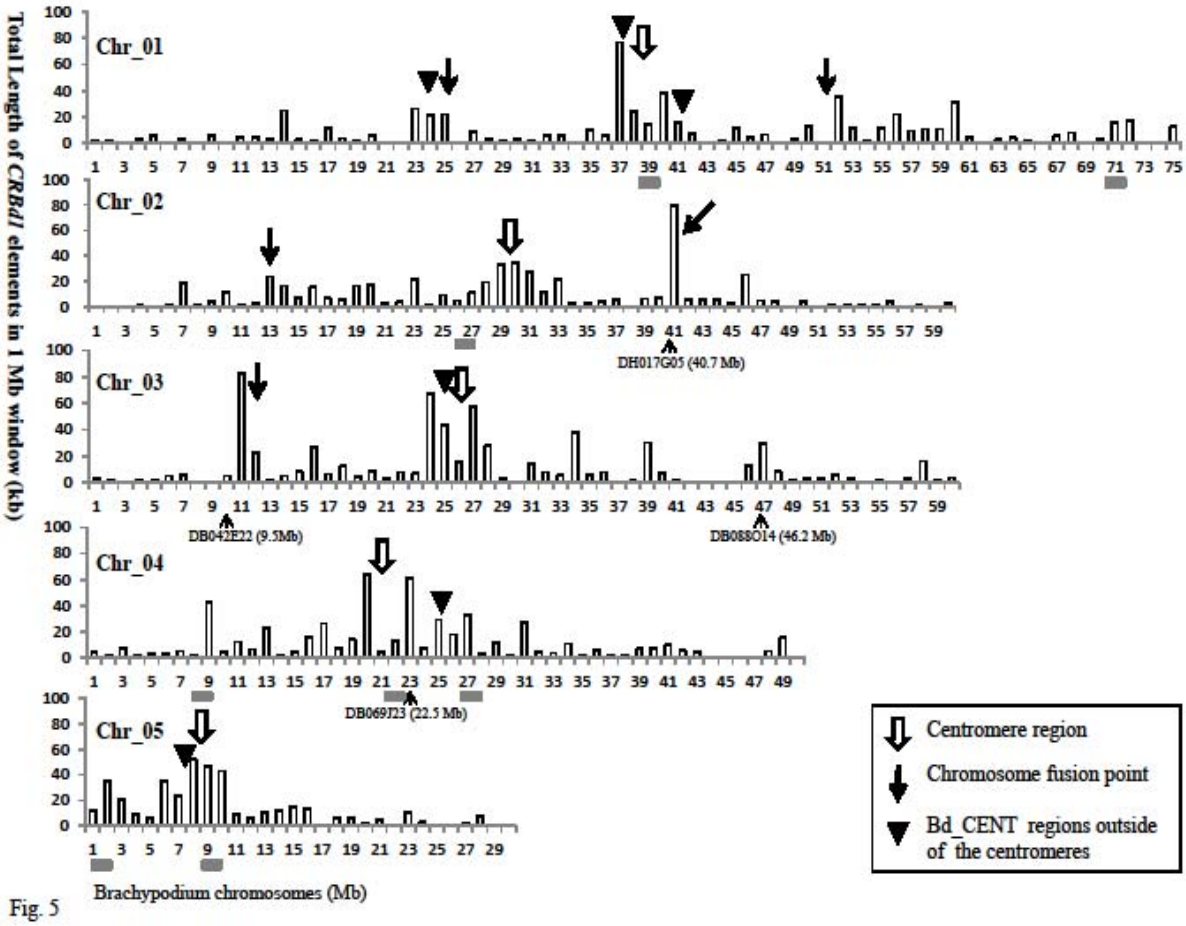

Fig. 5

Fig. 5 The distribution histograms of the LTR of *CRBd1* on the *Brachypodium* chromosomes. The total sequence length (kb) of alignments in 1Mb window was plotted along the chromosomes. The positions of centromeres, chromosome fusion points, and Bd_CENT containing regions outside of the centromeres were marked. The arrows point the positions of the four *Brachypodium* BACs, and the gray bars represent the positions of the eight full-length *CRBd1*.

Figure 6

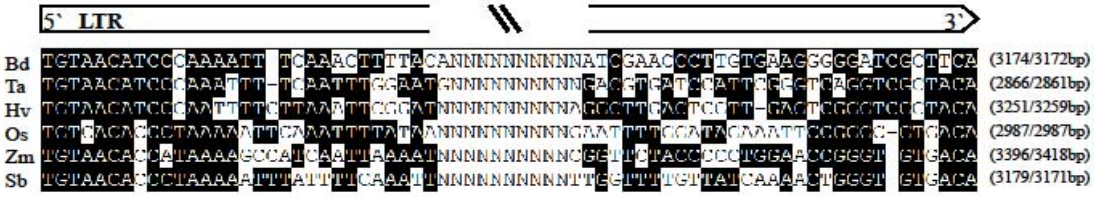

Fig. 6

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

1 Figure S1

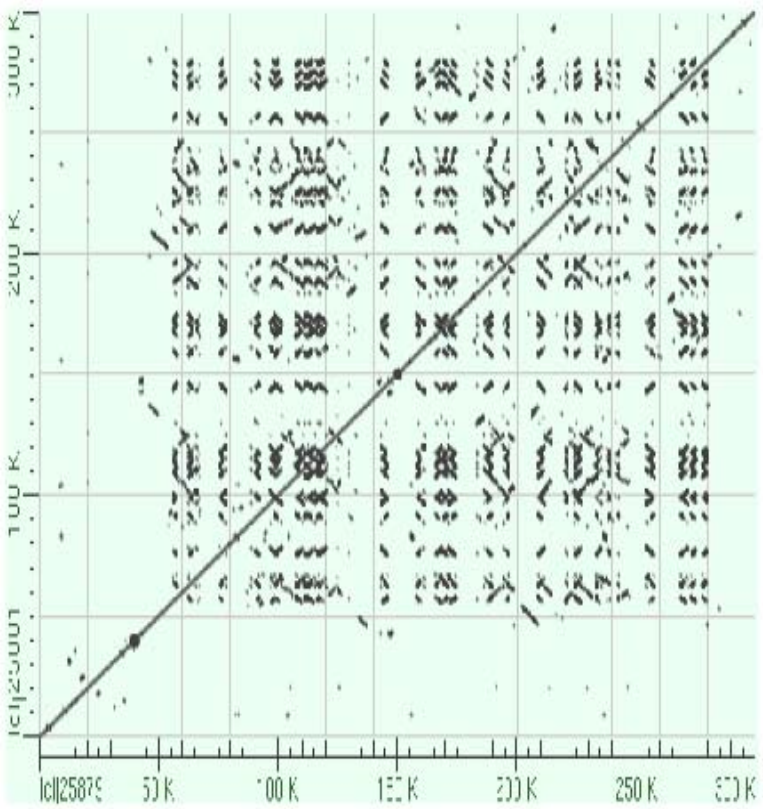

Fig. S1

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