Sequence organization of barley centromeres

Sabina Hudakova, Wolfgang Michalek, Gernot G. Presting, Rogier ten Hoopen, Karla dos Santos, Zuzana Jasencakova and Ingo Schubert*

Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Corrensstrasse 3, D-06466 Gatersleben, Germany

Received September 3, 2001; Revised October 22, 2001; Accepted November 1, 2001

DDBJ/EMBL/GenBank accession nos[†]

ABSTRACT

By sequencing, fingerprinting and *in situ* hybridization of a centromere-specific large insert clone (BAC 7), the sequence organization of centromeric DNA of barley could be elucidated. Within 23 kb, three copies of the *Ty3/gypsy*-like retroelement *cereba* were present. Two elements of ~7 kb, arranged in tandem, include long terminal repeats (LTRs) (~1 kb) similar to the rice centromeric retrotransposon RIRE 7 and to the *cereal centromeric sequence* family, the primer binding site, the complete polygene flanked by untranslated regions, as well as a polypurine tract 5' of the downstream LTR. The high density (~200 elements/centromere) and completeness of *cereba* elements and the absence of internally deleted elements and solo LTRs from the BAC 7 insert represent unique features of the barley centromeres as compared to those of other cereals. Obviously, the conserved *cereba* elements together with barley-specific G+C-rich satellite sequences constitute the major components of centromeric DNA in this species.

INTRODUCTION

The centromere of monocentric chromosomes is morphologically recognizable as the primary constriction. Centromeres are essential for correct segregation into daughter cells of sister chromatids during mitosis and meiosis II and of homologous chromosomes during meiosis I (reviewed in 1,2). Although centromere function is highly conserved among eukaryotes, as are the kinetochore proteins including those of higher plants (3,4), centromeric DNA is rather variable.

Centromeric DNA sequences have been described for several eukaryotes. However, except for some yeasts (5,6), their functional importance is at least controversial, the more so since for several species neocentromeric activities at non-centromeric positions have been reported, supporting the idea that the centromere location might be regulated epigenetically (7).

While for some plants no centromere-specific repeats could be isolated (8) such sequences have been found in others. For instance, the 180 bp repeat of *Arabidopsis* (9), which forms large tandem arrays with the repeat 106B (10) interspersed therein, occupy the central domain of all five *Arabidopsis* centromeres (11–13).

A few years ago, two centromeric sequences were described for cereals. One is the *cereal centromeric sequence* (CCS1) family of *Brachypodium* that also occurs in wheat, rye, barley, maize and rice centromeres (14); the other is the *Sau3A9* sequence of sorghum which also hybridized to the primary constriction of the above species (15). Thereafter, using a barley homolog of *Sau3A9* as a probe, a λ clone (#9) from a genomic library was detected containing a *cereba* (centromeric retroelement of barley) element with high similarity to the *Ty3/gypsy* group of retrotransposons (16). This element hybridized to all barley centromeres. It contained a complete polygene, of which *Sau3A9* represents the integrase encoding region, and flanking sequences similar to CCS1, supposed to represent long terminal repeats (LTRs) of *cereba*. Due to the additional presence of BARE retroelement sequences (dispersed along the chromosome arms of barley; 17) within the λ 9 clone and a *DraI* restriction pattern, which differed from that of genomic DNA when probed with the barley homolog of *Sau3A9*, we assumed that this clone might contain either sequences of a centromere-border or a chimeric insert not really representative of barley centromeres. Meanwhile, further conserved sequences representing parts of *gypsy*-like retroelements were found within the centromeres of several cereals [CentA in maize (18); pHind22 in sorghum, wheat, maize and rye (19); RCS1 in rice, rye, barley, sorghum and maize (20); RCB11 in rice and *crwydryn* in oats and rye (21,22); RIRE7 in rice (23–25); R11H in wheat (26)] and even of *Beta* species [pBv26 and pBp10 (27)]. However, completeness and arrangement of these sequences have not yet been studied directly by complete sequencing of large insert clones.

The aim of this work was to search for large insert clones harboring sequences representative of barley centromeres and to study the sequence organization of *cereba* elements and possibly associated centromere-specific sequences within the corresponding clone in comparison with sequences of other cereal centromeres. This should provide suitable candidate

*To whom correspondence should be addressed. Tel: +49 39482 5239; Fax: +49 39482 5137; Email: schubert@ipk-gatersleben.de

Present addresses:

Gernot G. Presting, Novartis Agriculture Discovery Institute, 3115 Marryfield Row, San Diego, CA 92121-U25, USA

Karla dos Santos, Institut für Pflanzenbau und Pflanzenzüchtung, Universität Göttingen, von Siebold Strasse 8, 37075 Göttingen, Germany

[†]AY040832, AY040833

Figure 1. FISH with BAC 7 as probe on barley chromosomes (A) and size determination of BAC 7 by PFGE (B). (A) BAC 7 labeled with rhodamin-5-dUTP yielded signals exclusively at the centromeres of all barley chromosomes of karyotype MK 14/2034; (B) BAC 7 DNA linearized by *XhoI* (left) and digested by *NotI* (right); *NotI* yielded a vector fragment (6.9 kb) and the insert (23 kb) flanked by minor parts of the vector. Bar represents 10 μ m.

sequences for gel-shift assays with kinetochoric proteins to characterize interactions between DNA and proteinaceous components of the barley centromere/kinetochore complex. We describe the insert sequence of a selected barley BAC clone which yielded similar hybridization patterns as genomic DNA using the barley centromeric sequence (BCS2) (14) and the integrase encoding region as probes and which hybridized to all barley centromeres *in situ*.

MATERIALS AND METHODS

BAC library screening and fluorescent *in situ* hybridization (FISH)

A BAC library of genomic DNA from *Hordeum vulgare* L. cultivar Morex (established at Clemson University) containing 313 344 clones (28) was transferred onto Hybond N⁺ filters (Amersham). Treatment of the filters, hybridization and washing conditions were as described (29,30). Of 10 BAC clones that hybridized with the integrase region (pGP7) of the polyprotein gene of Ty3/gypsy-like retrotransposon *cereba* labeled with [³²P]dCTP [using a random primer extension kit (Amersham) as described previously (31)] only one (03J24, now called BAC 7) showed after FISH positive signals exclusively at the centromeric regions of all barley chromosomes.

For FISH, BAC 7 DNA was isolated using a Qiagen Plasmid Mini Kit (100) and labeled with rhodamin-5-dUTP using a nick translation kit (Roche Biochemicals) according to the manufacturer's instructions. The primers (AGGGAG)₄ and (CTCCCT)₄, representing the most frequent motif within the G+C-rich domain outside the *cereba* elements of the BAC 7 insert, were amplified without additional template sequence and biotin labeled by PCR as described previously (32) for FISH.

Metaphase spreads from root tip meristems of the barley line MK 14/2034 (characterized by two homozygous reciprocal translocations between chromosomes 3H/4H and 7H/5H) were prepared as described (16), FISH and signal detection were performed as described (33).

BAC size determination

The size of the BAC 7 clone was measured by pulsed field gel electrophoresis (PFGE) using the CHEF-DR® II electrophoresis system (Bio-Rad) with a 5 s pulsed time (5 V/cm) for 15 h on a 1% agarose gel (Gibco BRL) at 14°C in 0.5× TBE buffer (45 mM Tris-borate, 1 mM EDTA, pH 8.0). A λ /HindIII ladder (MBI Fermentas) was used as a molecular weight marker.

Restriction digests, agarose gel electrophoresis and Southern blot analysis

For restriction analysis, aliquots containing 70 ng of BAC 7 DNA were completely digested for 3 h at 37°C with 10 different restriction endonucleases (*Bgl*II, *Bst*XI, *Eco*RI, *Hind*III, *Kpn*I, *Not*I, *Pst*I, *Sal*I, *Sfi*I, *Xba*I) and 20 double combinations. The digestion products and the molecular weight markers Smartladder (EUROGENTEC) and Gene Ruler™ DNA Ladder Mix (MBI Fermentas) were electrophoresed on 0.8% agarose gels (Gibco BRL, Life Technologies) in 1× TBE buffer at 78 V for 4 h.

To perform Southern blot analysis, single or double digests of BAC 7 DNA with the restriction enzymes *Eco*RI, *Hind*III, *Pst*I, *Not*I, *Sal*I were carried out. The fragments were separated on 1% agarose gels and blotted onto a Hybond-N⁺ nylon membrane (Amersham Life Science) in 20× SSC solution. The DNA was fixed on the membrane by exposure to UV light for 3 min. Prehybridization and hybridization were performed overnight at 68 and 58°C, respectively, in 5× SSC, 0.1% (w/v) N-lauroylsarcosine, Na-salt (Sigma), 0.02% (w/v) SDS and 0.5% (w/v) blocking reagent (Boehringer Mannheim). As probes we used pBeloBAC 11 (vector) and the following inserts of subclones of the λ 9 clone (accession number AF078801; 16), which represent parts of the retrotransposon *cereba*: pGP7 (1.5 kb, RNase H + integrase domain), pGP12 (1.6 kb, gag + RNA binding domain), pGP33 (1.6 kb, including 182 bp homologous to the barley variant of CCS1) (14), pGP5 (1.1 kb, reverse transcriptase domain) and pGP13 (0.46 kb, protease domain). The pGP inserts were obtained by digestion of the subclones pGP7 and pGP5 with *Xba*I and

Table 1. Fragments of BAC 7 after complete digestions with 10 individual restriction enzymes and 20 pairwise combinations

Bgl II	BstXI	EcoRI	HindIII	Kpn I	Not I	Pst I	Sal I	Sfu I	Xba I
6.2	8.0	9.2 12 33 v	7.5 V	9.4 ~23.0	~11.5	5 7 12 13 33	~22.0 5 7 12 13 33 v	~14.0	~11.3
4.4	5.6	8.0 5 7 12 13 33	4.8 *	6.0	6.9	4.3 12 33	6.4 V	7.0	8.0
3.9	4.8	6.6 V	4.1 d 5 7 13	5.6		3.8 5 7 12 13	0.82 V	4.5	5.3
3.5 d	4.3	5.0 5 7 12 13 33 v	2.9 t 12 33	3.27		3.4 12 v		2.4	2.7
2.4	3.5 d	0.87 7	0.9 *	2.15 d		2.9 V		1.3	2.4
2.0				1.15		2.5 12 33		0.87	
1.5 d						1.5 V			
0.3 d									
Σ 29.5	29.7	29.67	30.1	29.72	29.9	29.9	29.22	30.07	29.7
Bgl II	BstXI	BstXI	BstXI	BstXI	BstXI	EcoRI	EcoRI	EcoRI	EcoRI
HindIII	EcoRI	HindIII	Pst I	Sfu I	Xba I	HindIII	Not I	Pst I	Sal I
4.8	8.0	4.8	8.2	5.8	8.1	6.6 V	8.0 d 5 7 12 13 33	8.4 12 13	9.2 12 33 v
3.3	4.3	4.0	3.8	4.8	4.3	4.8 *	6.4 V	4.3 12 33	8.0 5 7 12 13 33
2.3	3.4	3.4	3.4	4.6	3.4 t	2.9 t 5 7 12 33	8.0 d 5 7 12 13 33	2.9 V	5.8 V
2.1 t	3.2	2.8 d	3.2 d	3.5 d	2.6	2.0 d 5 7 13	0.87 7	2.5 12 33 v	5.0 5 7 12 13 33
2.0	2.7 d	2.3	2.9	2.4 d	2.2	1.9 5 7 13	0.6 V	2.4 d 5 7 12 13	0.87 7
1.4 q	2.3	2.2 d	2.3	1.1	1.9	1.2 5		2.2 V	0.8 V
0.9 t	1.6	1.3 d	1.0	0.95		0.9 *		1.5 V	0.56 V
0.68 d	0.8 d	0.9	0.5 d			0.87 7		1.2 d 5 7 12 v	
0.4		0.7 t				0.8 V		0.87 7	
0.3									
Σ 29.06	29.8	29.4	29.0	29.05	29.3	29.77	26.87	29.87	30.23
EcoRI	EcoRI	HindIII	HindIII	HindIII	HindIII	Not I	Pst I	Pst I	Sfu I
Sfu I	Xba I	Not I	Pst I	Xba I	Pst I	Pst I	Sfu I	Xba I	Xba I
6.6	8.3	6.9 V	4.8 *	7.5	5.3	~11.3 5 7 12 13 33	4.8	~11.3	5.3 d
5.2	5.3 d	4.8 *	4.1 5 7 13	4.3	4.8	4.3 12 33	4.5	4.3	4.6
5.0	2.5 d	4.1 d 5 7 13	3.3 5 7 13	3.8 d	4.1 d	3.8 5 7 12 13	3.8	3.8	4.2 d
4.5	2.4	2.9 t 12 33	3.0 V	2.7 t	2.4 t	2.7 V	3.4	2.9	2.7
4.2	1.3 d	0.9 *	2.9 V	0.9 d	2.2	2.6 V	2.9	2.6	2.6
2.4	0.87	0.4 * (?)	2.5 t 12 33		0.9	2.5 12 33 v	2.7 d	2.4	2.3
1.8			1.5 V		0.5 d	1.5 V	2.5	1.5	
0.9			0.9 *			0.6 12 v	1.5	0.8	
0.87			0.78 *			0.32 * (?)	0.85		
			0.4 t 12						
Σ 31.47	29.77	29.9	29.98	29.3	29.6	29.58	29.65	29.6	31.2

Restriction fragments of BAC 7 were identified by Southern blot analyses with the vector, subclones of *cereba* representing the polygenic region of the retrotransposon [pGP5 (yellow), 7 (green), 12 (pink), 13 (orange)] and the supposed LTR [pGP33 (blue)].

Numbers, sizes of fragments in kb; bold, fragments representing only sequence of the vector pBeloBAC 11 (V); italics, fragments composed of vector (v) and insert sequences.

Where Southern blotting was not performed, the fragments corresponding either to the vector, or to the vector and insert sequences were determined by comparison with restriction fragments known for the vector.

*, fragments that hybridized neither with sequences of *cereba*, nor with the vector; d,t,q, double, triple, quadruple band; (?), fragments which should hybridize with the vector but did not.

HindIII and of pGP12, pGP13 and pGP33 with *EcoRI* and *HindIII* and extraction from gels using a QIAEX Kit (Qiagen). Probes were labeled using a Dig-high prime Kit (Boehringer Mannheim), according to the supplier's instructions. After hybridization, the membrane was washed twice in 2× SSC, 0.1% SDS for 5 min at room temperature and twice in 0.1× SSC, 0.1% SDS for 5 min at 58°C. The DNA–DNA hybrids were detected by chemiluminescence with the CSPD® Kit (Boehringer Mannheim). Prior to reuse, the membrane was stripped by boiling in 0.5% SDS.

Subcloning, shotgun sequencing and data analysis

BAC 7 DNA was sonicated and fragments (~550 bp) were subcloned into the pBluescript II SK- vector (Stratagene) and sequenced using an ALFexpress (Pharmacia Biotech) or an

ABI Prism 377 (Perkin Elmer) DNA sequencer. Of an ~4.8 kb *HindIII* fragment, ~3.9 kb flanked by G+C-rich domains could not be sequenced completely, even by two specialized Biotech companies. The shotgun-sequencing data were analyzed with the Sequencer 3.1.1. software (Gene Codes). The resulting contigs were compared with the GenBank entries for the λ9 clone and the CCS1 (position 1–260) of the Hi-10 clone derived from *B.sylvaticum* (U52217) at NCBI using the BLASTN homology search software (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>). The program TAIR BLAST™ Similarity Search (<http://arabidopsis.org/blast/>) was used for comparison of the BAC 7 insert sequence with other plant sequences of the GenBank database. The nucleotide sequence of BAC 7 clone has been deposited in

Figure 2. Sequence organization of the barley clone BAC 7. (A) Scheme of a gypsy-type retrotransposon; (B) scheme of the insert of $\lambda 9$ clone (15 230 bp); (C) organization of the insert of BAC 7 in comparison with $\lambda 9$; (D) the entire clone BAC 7; (D1) restriction map of the contig of 14 993 bp of the BAC 7 insert; (D2) restriction map of the contig of 3603 bp; (E) scheme of the first retrotransposon copy of the contig of 14 993 bp in comparison with other plant sequences (Table 3). Violet arrows, position and orientation of the LTRs; black bars, 3' and 5' untranslated regions; dark blue, homologous to parts of subclone of *cereba* (pG33) including the CCS1 domain; gag+RB, gag gene (encoding structural proteins involved in intracellular packaging of the RNA transcript) + RNA binding site; PR, protease; RT, reverse transcriptase; RH+I, RNase H + integrase; BARE, sequences containing regions of homology to a *copia*-like element of barley (17) and to related dispersed repeat elements of rye (37); PBS, primer binding site, hatched in blue; PPT, polypurine tract, hatched in green; dotted, non-sequenced region (~3.9 kb) flanked by G+C-rich domains of BAC 7; hatched in black, regions corresponding to insert of $\lambda 9$ outside of *cereba* sequences; white, sequence absent from $\lambda 9$, numbers below (C) indicate length in bp; arrows in (B), orientation of the polygene domain; colored bars in (C) specify regions of BAC 7 aligned with $\lambda 9$ (Table 2); numbers in (E) correspond to sequences specified in Table 3; numbers in parentheses next to enzymes (B = *Bgl*III, Bs = *Bst*XI, E = *Eco*RI, H = *Hind*III, K = *Kpn*I, P = *Pst*I, X = *Xba*I) indicate the positions of restriction sites in (D1) and (D2).

GenBank under the accession numbers AY040832 and AY040833.

RESULTS

Isolation and characterization of the centromere-specific clone BAC 7

A genomic barley BAC library was screened with pGP7, a plasmid subclone of the $\lambda 9$ clone, which is highly homologous to the integrase region of the polyprotein gene of the *Ty3/gypsy* group of retrotransposons. Ten clones were selected, but only one of these (BAC 03J24, later called BAC 7) showed a positive FISH signal exclusively at the centromeric regions of all barley chromosomes (Fig. 1A). The other nine BACs also yielded dispersed signals. They either contain additionally centromere-flanking sequences, e.g. *copia* elements, as is the case for $\lambda 9$ (16) and some large insert clones of maize and sorghum (18,19) which contain *gypsy*- as well as *copia*-like

elements, or represent chimeric inserts, or contain (parts of) centromeric sequences which may occur as single copies outside centromeres. FISH with genomic DNA of barley revealed uniformly dispersed signals along all chromosomes, while a 10-fold excess of unlabeled BAC 7 DNA suppressed signals at centromeres, suggesting that BAC 7 contains the major sequence components of barley centromeres. BAC 7 showed a hybridization pattern similar to that of genomic DNA after digestion with *Dra*I and Southern hybridization with pGP7 and BCS2, the barley homolog of the CCS1 family (14). These observations and the fact that all sequence components of BAC 7 are detectable by FISH exclusively at all centromeres (see below and 16) indicates that the insert of BAC 7 is indeed representative for barley centromeric sequences.

To determine the size of BAC 7, its DNA was isolated, digested with *Not*I and *Xho*I, respectively, and separated by PFGE (Fig. 1B). Digestion with *Xho*I yielded only one band corresponding to the linearized plasmid (~30 kb), while *Not*I

Figure 3. FISH with (AGGGAG)_n, amplified and biotin-labeled by PCR, yielded signals on all centromeres of metaphase chromosomes (A) and interphase nuclei (B) of barley karyotype MK14/2034. Bars represent 10 μm.

yielded two fragments, one (~6.9 kb) comprising most of the vector and the other one (~23 kb) the insert flanked by short stretches of vector DNA at both sides. Therefore, the insert size was estimated to be ~22 500 bp, which together with the vector pBeloBAC 11 (7507 bp) constitute BAC 7.

Sequencing and restriction fragment mapping of BAC 7

Shotgun sequencing of 150 subclones of BAC 7 with an average size of 550 bp was performed. Because the occurrence of repetitive sequences was to be expected for the insert of BAC 7 and sequencing and alignment into contigs is difficult for such sequences, a restriction map of BAC 7 was constructed in parallel. For that purpose, DNA of BAC 7 was completely digested with 10 restriction enzymes and 20 pairwise combinations and electrophoresed on agarose gels. Southern blots were hybridized consecutively with five subclones of λ9 as well as with the vector pBeloBAC 11 as probes. A compilation of the resulting fragments is given in Table 1. As expected, all fragments per digest amounted to ~30 kb, the size of the entire BAC 7. Double or triple bands were determined by comparing band intensities to that of the molecular weight markers. The restriction map has been designed manually by assembling restriction fragments from single and multiple digests in comparison with the sequence alignment obtained from shotgun sequencing data (Fig. 2D1 and D2). This led to mutual control and confirmation of data (sequence alignment versus fingerprinting). The entire insert of BAC 7 (~22 500 bp) revealed a contig of 14 993 bp, separated from a second contig of 3603 bp by a fragment of ~3900 bp, flanked on either side by G+C-rich sequences of 349 and 776 bp, respectively (Fig. 2D). The internal part of this fragment could not be sequenced completely by two specialized Biotech companies. Subclones of this fragment revealed mainly the motif AGGGAG and degenerated versions of it, but no new sequences. Tetrameres of the AGGGAG motif and its complementary sequence were used as primers for PCR with only nucleotides and *Taq* polymerase. The primer-multimer products yielded a smear on agarose gels and strong FISH signals exclusively at all centromeres of barley (Fig. 3) but not on rye and wheat centromeres.

Table 2. Sequence comparison (BLASTN) of BAC 7 and λ9

n	contig 1 (14,993 bp) (positions)	λ 9 (positions)	similarity [%]	identity [in bp]
1	33-262	2038-1815	80	187/231
2	104-262	11643-11489	83	132/159
3	253-833	2074-2645	84	496/589
4	898-973	2705-2780	89	68/76
5	1007-1639	2826-3464	91	582/639
6	1760-5980	3561-7777	94	3976/4227
7	6239-6277	8556-8517	90	36/40
8	6852-6896	7835-7791	88	40/45
9	6935-7176	2055-1820	81	198/243
10	7023-7160	11643-11508	83	115/138
11	7826-8071	2055-1815	81	201/247
12	7914-8071	11643-11489	82	131/158
13	8062-8536	2074-2542	85	409/481
14	8588-8642	2591-2645	92	51/55
15	8707-8782	2705-2780	89	68/76
16	8827-9459	2826-3464	90	581/693
17	9571-13791	3561-7777	94	3976/4227
18	14050-14088	8556-8517	90	36/40
19	14663-14707	7835-7791	88	40/45
20	14824-14977	1984-1834	81	122/155
21	14843-14977	11640-11508	78	110/135
contig 2 (3,603 bp)				
22	598-642	7835-7791	88	40/45
23	681-925	2055-1815	80	199/246
24	769-925	11643-11489	81	128/157
25	916-1488	2074-2645	85	496/581
26	1563-1638	2705-2780	88	67/76
27	1716-2348	2826-3464	90	580/639
28	2460-3603	3561-4698	90	1041/1144

From sequence comparison of BAC 7 with the components of *cereba* of λ9 a high degree of similarity became evident in spite of some rearrangements outside the polygene region (Table 2 and Fig. 2C). Within both contigs the RNA binding domains of the BAC 7 *cereba* elements show insertions of 119 bp (position 1640–1759) and 110 bp (position 9460–9570 and position 2349–2459 within the shorter contig, respectively) in addition to insertions within the 5' and 3' untranslated

Table 3. Comparison (BLASTN) of the first retrotransposon copy on contig of 14 993 bp (position 1–7176) of BAC 7 with other known plant centromeric sequences

similarity to	domain /accession/	position on the original accession	position on BAC 7	identity [%]
<i>O. sativa</i>	polyprotein region of the <i>gypsy</i> -like retrotransposon RIRE 7 /AB033235/; (24)	3187-5896	3545-6254 [1], 11356-14065	81
		2579-3375	2943-3375 [2], 10754-11186	72
		1884-2091	2248-2455 [3], 10059-10266	73
		1608-1662	1972-2026 [4], 9783-9837	81
<i>O. sativa</i> subsp. <i>indica</i>	dispersed centromeric repeat family RCS1 /AF078903/; (19,20)	27-1184	5119-6275 [5], 12930-14086	82
<i>O. sativa</i>	centromeric sequence RCB 11 /AB013613/; (21)	2227-2681	5128-5582 [6], 12939-13393	81
		3121-3372	6022-6273 [7], 13833-14084	83
<i>S. bicolor</i>	centromeric sequence pHind22 /AF078901/; (19)	20-366	5647-5993 [8], 13458-13804	82
<i>S. bicolor</i>	centromeric sequence pSau3A9 /SBU68165/; (15)	5-324	4996-5315 [9], 12807-13126	82
		338-422	6184-6268 [10], 13995-14079	90
<i>Z. mays</i>	centromeric retrotransposon-like repeat CentA /AF078917/; (18)	698-1071	3982-4355 [11], 11793-12166	79
		6162-6402	4810-5050 [12], 12621-12861	82
		5949-6094	4597-4742 [13], 12408-12553	82
		407-492	3691-3776 [14], 11502-11587	86

Numbers in parentheses indicate the position of sequences shown in Figure 2E.

regions between the LTRs and the polygene regions as compared to the corresponding sequence of *cereba* of $\lambda 9$ (Table 2 and Fig. 2C).

The contig of 14 993 bp contains two *cereba* elements in tandem

The left contig of the BAC 7 insert is formed by two almost identical and complete *Ty3/gypsy*-like retroelements of high similarity to *cereba* of $\lambda 9$. Both contained all five catalytic regions (RNA binding site/protease/reverse transcriptase/RNase H/integrase) including *gag*, primer binding sites (PBSs) and polypurine tracts (PPTs) and are flanked by LTRs on both sides. The upstream LTR of the first element (position 1–257) lacks the first 665 bp. The PBS follows immediately at its 3' end (position 258–274). The downstream LTR (922 bp) with a terminal TGAT/ATCA inverted repeat is preceded by a PPT at position 6242–6254 (Fig. 2D1). The second copy of the *cereba* element is complete except for the first 23 bp lacking at the upstream LTR. The 5'-regions of the (almost) complete LTRs show similarity (~50%) with the LTR sequence of the RIRE 7 *gypsy*-type retrotransposon in rice (24) and the last third of LTRs (260 bp) with the CCS1 sequence (80%).

Both retrotransposons show extended homology to the sequence RCB 11 (AB013613; 21), the *gypsy*-type retrotransposon RIRE 7 (AB033235; 24) and the repeat RCS 1 (AF078903; 19,20) of rice, to the repeats *pSau3A9* (SBU68165; 15) and *pHind22* (AF078901; 19) of sorghum, as well as to the retrotransposon-like repeat CentA of maize (AF078917; 18), which all occupy centromeric positions (Table 3 and Fig. 2E).

The contig of 3603 bp contains a truncated *cereba* element

The right contig of the BAC 7 insert between the G+C-rich stretch and the vector covers a *cereba* element extending from the 5' LTR (position 1–920) with the terminal TGAT/ATCA inverted repeat up to the end of the *gag*+RB region. This element is nearly identical to the corresponding parts of the complete *cereba* elements of the left contig.

DISCUSSION

We have shown that the *gypsy*-like retroelement *cereba* occurs in three copies within 23 kb and that apparently complete and autonomous *cereba* elements including LTRs of nearly 1 kb constitute together with the G+C-rich satellite sequences, the major DNA component of barley centromeres. The density of retroelements is higher in barley centromeres than calculated for wheat (one *gypsy*-like element per 55 kb; 26) or sorghum centromeres (two such elements within a 90 kb BAC; 19). According to a previous estimation of about 200 *cereba* elements per barley centromere (16), this would amount to at least 1.4 Mb of centromeric DNA per chromosome. Also the completeness of the *cereba* elements is a novelty when compared to those within centromeric clones of other cereals (22). The assumption that CCS1 sequences form (parts of) LTRs (16) could be confirmed. The sequence and restriction analyses of BAC 7 are in accordance with the previous assumption that the $\lambda 9$ insert is not representative for the sequence organization within barley centromeres, albeit large insert clones combining centromeric *gypsy*-like and non-centromeric *copia*-like elements were also reported for sorghum (19) and maize (18). These clones are either chimeric or originate from centromere-flanking regions. Obviously, related elements are conserved within the centromeres of all cereals since their radiation ~60 million years ago, because a horizontal transfer between the contemporary cereal species is rather improbable (35). Furthermore, it is suggested that these types of retrotransposons do not frequently invade non-centromeric positions within their host genomes. Probably, they represent favoured targets for kinetochore assembling. This may be tested in future by gel shift assays with suitable plant kinetochoric proteins.

The absence of internally deleted elements or solo LTRs from BAC 7 is in contrast to observations made for rice (20,24), maize (18) and sorghum (34). It might be possible that centromere-specific satellites such as the G+C-rich sequence motif of barley have originated during evolution by nested

transposition (36); their redundancy may depend on species- and position-specific transposition frequencies of certain types of mobile elements which may (22) or may not be identical with those found to be clustered at cereal centromeres. Provided there is indeed a trend to substitute the centromeric gypsy-like elements by satellite sequences (22), then barley represents an early stage of this development due to its high density of centromeric retroelements.

ACKNOWLEDGEMENTS

We thank Jörg Plieske for advice with PFGE, Stephan Streng for his help with the BAC library screening, Paul Fransz and Dagmar Schmidt for stimulating discussion, Barbara Hildebrandt for technical assistance and Rigomar Rieger for critical reading of the manuscript. This work was supported in part by a grant from the DFG (Schu 951/6-1).

REFERENCES

- Choo, K.H.A. (1997) *The Centromere*. Oxford University Press, Oxford, UK.
- Maney, T., Ginkel, L.M., Hunter, A.W. and Wordeman, L. (1999) The kinetochore of higher eucaryotes: a molecular view. *Int. Rev. Cytol.*, **194**, 67–131.
- Dawe, R.K., Reed, L.M., Yu, H.-G., Muszynski, M.G. and Hiatt, E.N. (1999) A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. *Plant Cell*, **11**, 1227–1238.
- ten Hoopen, R., Manteuffel, R., Dolezel, J., Malysheva, L. and Schubert, I. (2000) Evolutionary conservation of kinetochore protein sequences in plants. *Chromosoma*, **109**, 482–489.
- Clarke, L. and Carbon, J. (1985) Structure and function of yeast centromeres. *Annu. Rev. Genet.*, **19**, 29–56.
- Hieter, P., Pridmore, D., Hegemann, J., Thomas, M., Davis, R. and Phillipsen, P. (1985) Functional selection and analysis of yeast centromeric DNA. *Cell*, **42**, 913–921.
- Karpen, G.H. and Allshire, R.C. (1997) The case for epigenetic effects on centromere identity and function. *Trends Genet.*, **13**, 489–496.
- Houben, A., Brandes, A., Pich, U., Manteuffel, R. and Schubert, I. (1996) Molecular-cytogenetic characterization of a higher plant centromere/kinetochore complex. *Theor. Appl. Genet.*, **93**, 477–484.
- Martinez-Zapater, J.M., Estelle, M.A. and Somerville, C.R. (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol. Gen. Genet.*, **204**, 417–423.
- Pélissier, T., Tutois, S., Tourmente, S., Deragon, J.M. and Picard, G. (1996) DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in *Athila* retroelement sequences. *Genetica*, **97**, 141–151.
- Fransz, P., Armstrong, S., Alonso-Blanco, C., Fischer, T.C., Torres-Ruiz, R.A. and Jones, G. (1998) Cytogenetics for the model system *Arabidopsis thaliana*. *Plant J.*, **13**, 867–876.
- Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T. and Jones, G.H. (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell*, **100**, 367–376.
- Heslop-Harrison, J.S., Murata, M., Ogura, Y., Schwarzacher, T. and Motoyoshi, F. (1999) Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. *Plant Cell*, **11**, 31–42.
- Aragón-Alcaide, L., Miller, T., Schwarzacher, T., Reader, S. and Moore, G. (1996) A cereal centromeric sequence. *Chromosoma*, **105**, 261–268.
- Jiang, J., Nasuda, S., Dong, F., Scherrer, C.W., Woo, S.-S., Wing, R.A., Gill, B.S. and Ward, D.C. (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. *Proc. Natl Acad. Sci. USA*, **93**, 14210–14213.
- Presting, G.G., Malysheva, L., Fuchs, J. and Schubert, I. (1998) A *TY3/GYPSY* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *Plant J.*, **16**, 721–728.
- Manninen, I. and Schulman, A.H. (1992) Analysis of BARE-1, a retrotransposon in barley. *Physiol. Plantarum*, **85**, A6 33.
- Ananiev, E.V., Phillips, R.L. and Rines, H.W. (1998) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc. Natl Acad. Sci. USA*, **95**, 13073–13078.
- Miller, J.T., Dong, F., Jackson, S.A., Song, J. and Jiang, J. (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. *Genetics*, **150**, 1615–1623.
- Dong, F., Miller, J.T., Jackson, S.A., Wang, G.-L., Ronald, P.C. and Jiang, J. (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proc. Natl Acad. Sci. USA*, **95**, 8135–8140.
- Nonomura, K.-I. and Kurata, N. (1999) Organization of the 1.9-kb repeat unit RCE1 in the centromeric region of rice chromosomes. *Mol. Gen. Genet.*, **261**, 1–10.
- Langdon, T., Seago, C., Mende, M., Leggett, M., Thomas, H., Forster, J.W., Thomas, H., Jones, R.N. and Jenkins, G. (2000) Retrotransposon evolution in diverse plant genomes. *Genetics*, **156**, 313–325.
- Kumekawa, N., Ohtsubo, E. and Ohtsubo, H. (1999) Identification and phylogenetic analysis of gypsy-type retrotransposons in the plant kingdom. *Genes Genet. Syst.*, **74**, 299–307.
- Kumekawa, N., Ohtsubo, E., Ohtsubo, H., Ohmido, N. and Fukui, K. (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.
- Nonomura, K. and Kurata, N. (2001) The centromere composition of multiple repetitive sequences on rice chromosome 5. *Chromosoma*, **110**, 284–291.
- Fukui, K.-N., Suzuki, G., Lagudah, E.S., Rahman, S., Appels, R., Yamamoto, M. and Mukai, Y. (2001) Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat. *Plant Cell Physiol.*, **42**, 189–196.
- Gindullis, F., Desel, C., Galasso, I. and Schmidt, T. (2001) The large-scale organization of the centromeric region in *Beta* species. *Genome Res.*, **11**, 253–265.
- Yu, Y., Tomkins, J.P., Waugh, R., Frisch, D.A., Kudrna, D., Kleinhofs, A., Brueggeman, R.S., Muehlbauer, G.J., Wise, R.P. and Wing, R.A. (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. *Theor. Appl. Genet.*, **101**, 1093–1099.
- Hoheisel, J.D., Maier, E., Mott, R., McCarthy, L., Grigoriev, A.V., Schalkwyk, L.C., Nizetic, D., Francis, F. and Lehrach, H. (1993) High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell*, **73**, 109–120.
- Nizetic, D., Drmanac, R. and Lehrach, H. (1991) An improved bacterial colony lysis procedure enables direct DNA hybridization using short (10, 11 bases) oligonucleotides to cosmids. *Nucleic Acids Res.*, **19**, 182.
- Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**, 6–13.
- Ijdo, J.W., Wells, R.A., Baldini, A. and Reeders, S.T. (1991) Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *Nucleic Acids Res.*, **19**, 4780.
- Fransz, P.F., Alonso-Blanco, C., Liharska, T.B., Peeters, A.J.M., Zabel, P. and de Jong, J.H. (1996) High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. *Plant J.*, **9**, 421–430.
- Miller, J.T., Jackson, S.A., Nasuda, S., Gill, B.S., Wing, R.A. and Jiang, J. (1998) Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor*. *Theor. Appl. Genet.*, **96**, 832–839.
- Kumar, A. and Bennetzen, J.L. (1999) Plant retrotransposons. *Annu. Rev. Genet.*, **33**, 479–532.
- SanMiguel, P., Gaut, B.S., Tikhonov, A., Nakajima, Y. and Bennetzen, J.L. (1998) The paleontology of intergene retrotransposons of maize. *Nature Genet.*, **20**, 43–45.
- Rogowsky, P.M., Shepherd, K.W. and Langridge, P. (1992) Polymerase chain reaction based mapping of rye involving repeated DNA sequences. *Genome*, **35**, 621–626.