Composition and location of wheat BAC sequences marked by *Aegilops speltoides* subtelo-meric repeats

Salina EA[1], Sergeeva EM[1], Adonina IG[1], Shcherban AB[1], Belcram H[2], Huneau C[2], Chalhoub B[2]
[1] Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090, Russia, e-mail: salina@bionet.nsc.ru, [2]Organization and evolution of plant genomes, Unité de Recherche en Génomique végétale (URGV), Evry, France.

INTRODUCTION

Two regions are distinguished in the chromosomal end structure, the telomeric region that caps the chromosome tip and the adjacent subtelomeric region. Taken together the results of recent analysis of subtelomeric DNA and current views have allowed the subtelomeric region to be described as located in the distal chromosome region between telomeric and unique chromosome-specific DNA sequences. Approaches targeted to cloning and analysis of distinct repetitive DNA families located at the chromosomal ends have been widely used in analysis of subtelomeric DNA of plant (Salina et al., 2006). Use of degenerate telomeric primers and the Vectorette PCR approach has made it possible to isolate and map the DNA sequences adjacent to telomeric repeats (Kilian & Kleinhofs, 1992). Large-scale sequencing of the genomes of a number of organisms enabled the complete characterization of the structural organization of the plant subtelomeric regions in certain rice (Yang et al., 2005) and Arabidopsis chromosomes (Kou et al., 2006).

The construction of BAC libraries of *Triticum aestivum* has provided opportunities for examining extensive subtelomeric DNA regions of the wheat genome (BBAADD) (Chalhoub et al., 2004). Our relevant finding was that the tandem repeats Spelt1 and Spelt52 are specific markers of the chromosomal ends, mainly of the B and G genomes of polyploid wheats and their putative donor *Ae. speltoides* (Salina et al., 2006). This prompted the interesting possibility of using these repeats as probes for the identification of BAC subtelomeric clones of wheat. The aim of the current study was to isolate and examine wheat BAC sequences marked by Spelt1 and Spelt52 repeats.

MATERIALS AND METHODS

The BAC clones carrying subtelomeric tandem repeats Spelt1 and Spelt52 were obtained from a genomic BAC-library of *T. aestivum* cv Renan (Chalhoub et al., 2004). We used the following protocol for general analysis of selected BAC clones: (a) determination of insert lengths; (b) estimation of core content of repetitive sequence in cloned DNA fragment, including various repeats occurring in the subtelomeric regions in wheat, Spelt1, Spelt52, pSc119.2, pAs1, and the telomeric repeat; (c) *in situ* hybridization of BAC clones to mitotic metaphase chromosomes of common wheat.

Two telomere-associated tandem sequences Spelt1 and Spelt52, isolated from *Ae. speltoides*; pSc119.2 from *Secale cereale*; and pAs1 isolated from *Ae. tauschii* were used as probes (Salina et al. 2006; McIntyre et al. 1990, Nagaki et al. 1995). A telomeric repeat synthetic probe was obtained by PCR.

The complete nucleotide sequence was determined for BAC-clone 2050O8 from the hexaploid wheat cultivar Renan, with subtelomeric *in situ* localization. The sequencing was conducted by means of shotgun library sequencing and sequence assembly as described by Chantret et al., (2005). Sequence annotation was done according to guideline for Annotating wheat genomic sequences from the International Genome Wheat Sequencing Consortium (http://www.wheatgenome.org/pdf/wheat_gene_annotation_Release1-1.pdf). FISH was conducted according to Salina et al. (2006). Spelt52_2050O8 was isolated from BAC_2050O8 and used as a probe. To study the distribution among the diploid and polyploid wheat species of CACTA transposable element *Caspar*_2050O8, we used BAC_2050O8, as well as cas1 (CTG-2) and cas2 (transposase domain) derived from *Caspar*_2050O8 as probes. The total number of analyzed metaphases was 5-10 from 3 individual plants for each probe. To identify chromosomes carrying signals, we used probe combinations pSc119.2+BAC and pAs1+BAC. The chromosomes were arranged according to standard genetic nomenclature (Badaeva et al., 1996).

RESULTS AND DISCUSSION

Spelt1 and Spelt52 as probes for subtelomeric clone isolation. Use of Spelt1 and Spelt52 for screening of the BAC library of the 'Renan' hexaploid wheat cultivar enabled us to choose 9 clones presumably associated with the subtelomeric regions (Table 1). The choice of Spelt52-containing BAC clones was based on a PCR ladder pattern characteristic of tandem repeats. To choose Spelt1-containing BAC clones, we resorted to Southern hybridization.

BAC_2050O8 was shown to localize predominantly to the subtelomeric chromosome regions both in the entire clone and its distinct regions according to *in situ* hybridization. *In situ* hybridization with wheat BAC clones can effectively map repetitive DNA sequences (Zang et al., 2004). With this in mind, we attempted to isolate the BAC clones from long repeat-poor DNA regions. We succeeded in isolating fragments of about 30 kb and 40 kb from BAC clones 2322J20 and 478L20, respectively. These fragments were shown to be localized to subtelomeric chromosome regions.

The presence of Spelt52 tandem repeats with total length of 8500-10000 bp (the monomeric unit of 380-390 bp) in BAC clones of cv. Renan makes it advantageous as a

probe for BAC localization (Table1). We then performed *in situ* hybridization of a BAC clone 2050O8 PCR fragment (Spelt52_2050O8) on the Renan chromosomes. Spelt52_2050O8 hybridized at the end of chromosome arms 1BS, 4BL, 7BL. Based on *in situ* hybridization of the Spelt52 probe, there was good reason for believing that virtually all the BAC clones containing the Spelt52 repeats in more than 23 copies (8740 bp) are from the subtelomeric regions of one of the chromosomes 1BS, 4BL or 7BL.

The Spelt1-containing BAC clones, 2322J20 and 2424P01, partially overlapping each other, were definitely at subtelomeric locations. The subtelomeric localization of 2 clones, namely 2383A24 and 112D20, of the 9 studied in the current work was questionable because of low Spelt1 and Spelt52 occurrence and lacking support by the applied *in situ* hybridization approach.

Table 1. The results of comparative BAC-clone analysis

BAC clone	Selected with	Length kbp	Copy number**		Content of repeats %***	Chromosomal localization
			Spelt1 178 bp	Spelt52 380 bp		
2322J20	Spelt1	93	7	-	54	*2Subtelomeric
2383A24		113.5	6	-	80	Dispersed
2424P01		97	5	-	70	*4Subtelomeric
112D20		86.5	1-2	5	62	Dispersed
110O07	Spelt52	162.5	-	24	56	*3Subtelomeric
110B21		168	-	23	73	*3Subtelomeric
2050O8		120	-	27	25	*1Subtelomeric
478L20		153.5	-	27	33	*2Subtelomeric
1487N20		165.5	-	28	62	*3Subtelomeric

* estimated by *in situ* hybridization of the BAC clone[1], of long fragments[2], of Spelt52_2050O8[3]

*[4] estimated by overlapping with subtelomeric BAC

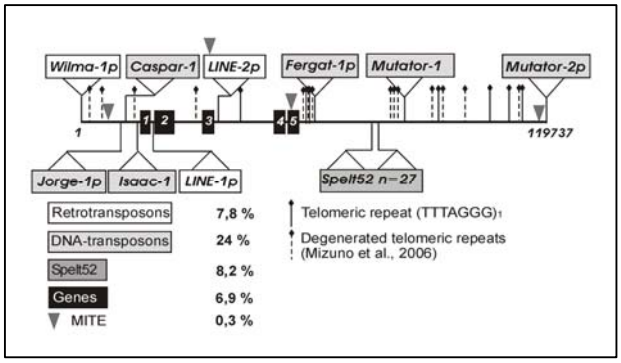
** estimated by Southern- and dot-blot hybridization methods, the monomer length is specified for each repeat family

*** estimated by Southern-blot hybridization method

Thus, the satellite DNA sequences from the subtelomeric regions of diploid progenitors can be efficiently used for choosing BAC clones from the corresponding regions of common wheat chromosomes.

Analysis of nucleotide sequence of subtelomeric BAC-clone 2050O8. For a more detailed analysis of subtelomeric DNA, the complete nucleotide sequence was determined for BAC_2050O8 with subtelomeric *in situ* localization. The general structure of 119.737 kb subtelomeric BAC-clone is shown in Figure1. The sequence consists of about 31.8% of transposable elements (TE), 17.1% Spelt52 and other non-TE-repeats and 6.9% non-TE-related genes. DNA transposons are predominant and make up 24.0% of the entire BAC-clone, whereas retrotransposons account for only 7.8% of the clone length. Full-length CACTA transposon Caspar covers 11,667 bp, encoding a transposase and CTG-2 proteins, and alone accounts for 54.6% of all transposable elements. Five hypothetical genes are identified as located in gene-containing region. The first two genes are homologous to the hypothetical rice genes and contain conserved domains of xyloglucan endotransglycosylase (XET) and riboflavin-deaminase-reductase, the third is a putative peroxisomal membrane protein PEX11-1. The others two contain hypothetical ORFs; one contains the Armadillo/beta-catenin-like repeats, and the last one is of unknown function predicted by FGENESH and GeneMark only.

Figure 1. Structural organization of *T. aestivum* subtelomeric BAC-clone 2050O8. In the names of the transposable elements, P at the end means a partial element with deletions of the portion of the ends. The proportion (%) of the different classes of transposable elements and other sequences are also presented.



BLAST alignments between BAC 2050O8 sequence and contigs containing mapped wheat ESTs (http://wheat.pw.usda.gov/GG2/blast.shtml) identified EST Contig 11876 (length 826 bp, location on 5AL, 4BL, and 4DL) with high homology (97%) in the 805 bp stretch to 2050O8 sequence and the region of homology is situated in the region of the putative peroxisomal membrane protein PEX11-1. It is of interest that comparison of the exon-intron structure between PEX11-1_2050O8 and EST Contig revealed a complete correspondence of the EST and the coding part of the gene. Analysis of restriction sites of BAC_2050O8 with mapping data obtained for EST Contig (available online at http://wheat.pw.usda.gov/cgi-bin/westsql/map_image.cgi) established the telomeric part of chromosome 4BL (bin 0.95 -1.0) as the location site of BAC_2050O8.

The long arm of chromosome 4B is known to be syntenic to the short arm of rice chromosome 3 (The rice chromosome 3 sequencing consortium, 2005). We performed a BLAST search of predicted genes of BAC_2050O8 and their protein products to the TIGR Rice genome annotation databases. We found that four predicted wheat genes (1-4, Fig.1), which showed significant homology to four putative rice genes, belong to OSJNBa0056G13 pseudomolecule (virtual contig) located in the distal part of rice chromosome 3S (http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice/name=OSJN Ba0056G13). These genes are located close to each other and their order is the same as that for the corresponding genes in BAC_2050O8, but one of the genes, namely Os03g02610, is in the opposite orientation in comparison with BAC_2050O8. The rice region enclosing these 4 genes is of 9556 bp, whereas the corresponding region in 2050O8 is 29612 bp. Such a difference is due to an enlarged intergenic region in 2050O8 (total length 22769 bp) against 3641 bp in rice. At the nucleotide sequence level the intergenic regions

of rice and wheat showed no similarity to each other. The high divergence level in the intergenic regions have been noted in almost all studies concerned with orthologous loci not only between wheat and rice (SanMiguel et al., 2002), but also between wheat species (Wicker et al., 2003, Chantret et al. 2005).

Caspar - transposable element of subtelomeric chromosomal regions of wheat and their diploid progenitors. To study the distribution among the diploid and polyploid wheat species of the CACTA *Caspar*_2050O8, transposable element, we used BAC_2050O8, as well as cas1 and cas2 derived from DNA regions of CTG-2 and transposase proteins as probes. Polyploid species of two evolutionary lines of wheat and their diploid progenitors were included in the analysis (Table 2). There was in general a good agreement between the locations of cas1, cas2 and BAC_2050O8 on the chromosomes of the studied cereals. All probes hybridized to subtelomeric chromosomal regions. Intraspecific polymorphism was observed in analysis of the distribution of probes on the chromosomes of *T. aestivum* cultivars. There were changes in signal intensity on chromosomes 1AL and 5AS in cvs Renan relative to CS and Saratovskaya 29.

Table 2. BAC_205008, cas2 and cas1 localization to subtelomeric chromosomal region of *Triticum* poplyploids and their diploid progenitors.

Species	**Probes localization to haploid genome.** Number of chromoso-mal regions (number of chromosomes)
Ae. speltoides	12 (7)
Ae. tauschii	5 (4)
T. urartu	13 (7)
T. monoccocum	13 (7)
T. durum	26-27 (14)
T. timopheevii	26-27 (14)
T. aestivum	
Chinese Spring	29 (18)
Renan	28 (17)
Saratovskaya 29	28 (17)

Futhermore, a hybridization signal was seen on the long arm of chromosome 6D in CS, while it was absent in the other studied cultivars. There were no hybridization sites on three chromosome pairs of hexaploid wheat when cas1, cas2 or BAC_2050O8 were used as probes; these chromosomes belong to the D genome. In tetraploids, *Caspar*_2050O8 probes were localized on all of the chromosomes, mainly in the subtelomeric regions of both arms. It is of interest that *T. durum*, like the *T aestivum*, chromosomes gave multiple strong signals of BAC_2050O8 hybridization to the distal third part of chromosome 4AL. The distribution pattern of the cas1, cas2, and BAC_2050O8, as probes among the *T. timopheevii* chromosome was uniform.

Interspecific differences were revealed in localization of the *Caspar*_2050O8 probes on the diploid species chromosomes. The number of hybridization sites varied from 12-13 in *Ae. speltoides* and species with the A genome, while their number was 5 in *Ae. tauschii*, and identified in 4 of the 7 chromosomes of the haploid genome (Table 2). Notice that the *Caspar* elements in BAC_A6 from *Ae. tauschii* localized also presumably to the subtelomeric chromosome regions of *Triticum* and *Aegilops* species (Zang et al., 2004). The *Caspar* of *Ae. tauschii* hybridized strongly to all wheat and *Ae. tauschii* chromosomes in contrast to *Caspar*_2050O8 of B genome, which is missing in 3 chromosomes of the D genome of common wheat and in 3 chromosomes of its progenitor *Ae. tauschii* (Table 2). Thus, the diverse modifications of *Caspar* during wheat evolution gave rise to genome-specific versions.

Li et al. (2004) suggested that the increase in copy number of *Caspar* elements toward the telomeres may be related to the high recombination rate in the distal regions. In this context of interest is the strongest hybridization of *Caspar*_2050O8 in the translocation region of *T. durum* and *T. aestivum* involving chromosome arms 4AL, 5AL, and 7BS.

Taken together, the evidence indicates that *Caspar* is a mobile element predominately in the subtelomeric regions of cereals and may be implicated in recombination in the distal chromosomal regions.

ACKNOWLEDGEMENTS

The work was supported by the program of "Dynamics of Genetics Resources and Biodiversity", by a grant from the Russian Foundation for Basic Research, project No. 08-04-00307.

REFERENCES

Badaeva, E.D., Friebe, B., Gill, B.S. (1996) *Genome*, **39(2)**, 293-306.

Chalhoub, B., Belcram, H., Caboche, M. (2004) *Plant Biotechnol J.*, 2(3), 181-188.

Chantret, N., Salse, J., Sabot, F. et al. (2005) *Plant Cell*, 17, 1033-1045

Kilian, A., Kleinhofs, A. (1992) *Mol. Gen. Genet.*, **235**, 153–156.

Kuo, H.F., Olsen, K.M., and Richards E.J. (2006) *Genetics*, **173(1)**, 401–417.

Li, W., Zhang, P., Fellers, J.P., Friebe, B. and Gill, B.S. (2004) *Plant J.,* 40, 500–511

McIntyre, C.L., Pereira, S., Moran, L.B., and Appels, R. (1990) *Genome*, 33, 635-640.

Nagaki, K., Tsujimoto, H., Isono, K., and Sasakuma, T. (1995) *Genome*, 38, 479-486

Salina, E.A., Lim, Y.K., Badaeva, E.D. et al. (2006) *Genome*, 49, 1023-1035

SanMiguel, P., RamaKrishna, W., Bennetzen, J.L., Busso, C., Dubovsky, J. (2002) *Funct. Integr. Genomics*, 2: 70–80

Wicker, T., Yahiaoui, N., Guyot, R., Schlagenhauf, E., Liu, Z.-D., Dubcovsky, J., Keller, B. (2003) *Plant Cell*, 15, 1186–1197.

Yang, T.J., Yu, Y., Chang, S.B. et al. (2005) *Theor. Appl. Genet.*, 111(3), 467-478.

Zhang, P., Li, W., Fellers, F.P. and Gill, B.S. (2004) *Chromosoma,* 112, 288–299.