Title: Effective Isolation of Retrotransposons and Repetitive DNA Families from the Wheat Genome

Running title: Effective Isolation of Retrotransposons and Repetitive Families

Authors: Motonori Tomita*, Munenori Asao, Aya Kuraki

Molecular Genetics Laboratory, Faculty of Agriculture, Tottori University, Tottori 680-8553, Japan

*Author for correspondence.

Tel: +81(0)857 31 5351;
Fax: +81(0)857 31 5351;
E-mail: <tomita206@gmail.com>, <tomita@muses.tottori-u.ac.jp>

Supported by Grants-in-Aid for Scientific Research (No. 01760004 and No. 04760006) from the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT).
Abstract

New classes of repetitive DNA elements were effectively identified by isolating small fragments of the elements from the wheat genome. A wheat A genome library was constructed from *Triticum monococcum* by degenerate cleavage with *EcoO109I*, the recognition sites of which consisted of 5′-PuGGNCCPy-3′ multi-sequences. Three novel repetitive sequences pTm6, pTm69 and pTm58 derived from the A genome were screened and tested for high copy number using a blotting approach. pTm6 showed identity with integrase domains of the barley Ty1-*Copia*-retrotransposon *BARE*-1 and pTm58 showed similarity to the barley Ty3-gypsy-like retrotransposon *Romani*. pTm69, however, constituted a tandem array with useful genomic specificities, but did not share any identity with known repetitive elements. This study also sought to isolate wheat D-genome-specific repetitive elements regardless of the level of methylation, by genomic subtraction. Total genomic DNA of *Aegilops tauschii* was cleaved into short fragments with a methylation-insensitive 4 bp cutter, *Mb*o*I*, and then common DNA sequences between *Ae. tauschii* and *T. turgidum* were subtracted by annealing with excess *T. turgidum* genomic DNA. The D genome repetitive sequence pAt1 was isolated and used to identify an additional novel repetitive sequence family from wheat BACs with a size range of 1395–1850 bp.

(198 words)

Keywords: *EcoO109I*; genome-specific DNA element; genomic subtraction; retrotransposon; tandem repetitive sequence family; wheat
There are several major cereal species grown for human consumption and a number, including wheat, barley, oat, rye and maize, have large genomes. Wheat (Triticum aestivum L.) is one of the principal caloric sources in the human diet. The wheat genome has the highest 1C DNA content of all Triticeae species (16.7 Gb; BENNET and LEITCH 2003). C₀t analyses, which measure the reassociation kinetics of single stranded nuclear DNA, have estimated that 75% of the wheat genome comprises repetitive sequences (FLAVELL et al. 1977). Of these, 16% differentiated to a wheat-specific array (RIMPAU et al. 1978). Rye (Secale cereale L.), a close relative of wheat, is a useful genetic resource for genes that confer biotic and abiotic stress resistance on wheat. The first man-made hybrid crop triticale was established by remote crossing between wheat and rye. Several repetitive sequences specific to rye have been identified from relic DNA fractions, and encode high–molecular weight sequences following methylation-sensitive cleavage. One particular family, known as the 350-family (BEDBROOK et al. 1980; CUADRADO et al. 1995; CUADRADO and JOUVE 1997) has been frequently used to analyze chromosome organization and evolution (ALKHIMOVA et al. 2004; CUADRADO and JOUVE 2002; YUAN and TOMITA 2009), and also to identify rye genomic components in wheat-rye crossbreeding (APPELS et al. 1982; LAPITAN et al. 1986; CUADRADO et al. 1997; FRANCKI et al. 1997; LIMA-BRITO et al. 1997). Although a number of wheat-specific repetitive sequences have also been identified (METZLAFF et al. 1986; RAYBURN and GILL 1986; TALBERT and CLARCK 1991; VERSHININ et al. 1994; UENG et al. 2000), only a single family, the pAs1 family (RAYBURN and GILL 1986) of Aegilops tauschii, has been successfully utilized in breeding (MUKAI et al. 1993; PEDersen and LANGRIDGE 1997; WANG et al. 2005; WANG et al. 2009). It therefore remains a goal to identify and exploit novel repetitive wheat clones for tagging the vast wheat genome.

As high-throughput sequencing of eukaryotic genomes has increased over recent years, interest in the structure and function of repetitive genomic sequences, previously referred to as junk DNA, has been revived. This has lead to a greater understanding of the amplification and rearrangement of repetitive sequences,
including transposable elements, which are now believed to play a significant role in genomic differentiation and evolution (Wendel 2000; Bennetzen et al. 2005). In addition, a substantial proportion of the genome is expressed as regulatory noncoding RNAs, some of which are reconstructed from transposable elements (Feschotte 2008; Brown et al. 2008; Siomi and Siomi 2008). In spite of advances in genome sequencing and annotation, the vast DNA databases that exists for higher eukaryotes still lack a majority of the genomic repetitive and noncoding sequences (Sabot et al. 2005; Stein 2007; Wicker et al. 2008), apart from homologues of known repetitive-sequence-family transposable elements or their derivatives (Flavell et al. 1992, 1994; Okada 1991; Kumar 1996; Kumar et al. 1997). In the wheat genome, the major constituent is repetitive arrays. Dozens of known transposable elements have been described, including Ty1-copia elements (Lucas et al. 1992), Ty3-gypsy elements, LINE elements, SINE elements, CACTA elements, and MITE elements, all of which have been registered in the GenBank database (http://www.ncbi.nlm.nih.gov/Genbank/index.html). Two unique types of repetitive elements, Revolver and Superior, have recently been identified in the rye genome through a process of genomic subtraction or by digestion with the restriction enzyme EcoO109I (Tomita et al. 2008, 2009b).

Our main goal is to identify novel repetitive components in the large, complex wheat genome, which might be useful tools in comprehensive genomics or molecular breeding. The wheat genome sequencing project has already been launched; however, progress is slowed by the large size and complexity of the wheat genome. Effectively targeting unknown repetitive sequences would facilitate greater efficiency in molecular breeding. In a previous study (Tomita et al. 2009a), we found that EcoO109I recognition sites, which consist of 5’-PuGGNCCPy-3’ multi-sequences, are present at a high frequency in repetitive families and recognized the potential for using this as a tool for studying genomic organization and differentiation. In the present study, we used EcoO109I and a genomic subtraction method to obtain repetitive elements from the wheat A and D genomes and successfully isolated a novel class of repetitive DNA elements.
Results

Isolation of A-genome-specific repetitive elements

The *Eco*O109I restriction enzyme recognizes sites consisting of 5'-PuGGNCCPy-3' multi-sequences, which are present at a high frequency in rye repetitive families (Tomita et al. 2009a). This enzyme was used to recover repetitive DNA elements from the *T. monococcum* genome. The genomic DNA of *T. monococcum* was digested completely with *Eco*O109I and the ends blunted. From the resulting elements, 1,000 ng was used for ligation and an equal amount of pUC19. By introducing this vector into *E. coli* JM109, we established a DNA library from the *T. monococcum* genome. Two hundred and seventy-three recombinant clones were randomly isolated and dot hybridization was performed with the total DNA of *T. monococcum*. The dot hybridization revealed that 21 of the 273 recombinant clones had strong hybridization signals. These clones represent highly repetitive clones with a copy number greater than 1,000. After agarose gel electrophoresis, the insertion size of the repetitive clone in the library was determined to be around 500 bp. The size of the clones might have been restricted by the cloning capacity of the pUC19 plasmid.

The core units of the screened repetitive clones were analyzed by Southern blot hybridization with *T. monococcum*, *T. turgidum*, *Ae. tauschii* and *T. aestivum* genomic DNA digested with *Eco*O109I. Southern analysis confirmed the hybridization specificity of eight of the 21 clones for *T. monococcum* (AA), *T. durum* (AABB) and *T. aestivum* (AABBDD), which all contain the A genome, but not for genomic DNA of *Ae. tauschii* (DD) and *S. cereale* (RR). The other 13 clones were non-genome-specific clones like pTm5 (Figure 1D), which showed smear hybridization patterns in all five genomes including *Ae. tauschii* (DD) and *S. cereale* (RR). The nucleotide sequences were determined for the clones pTm69, pTm6, and pTm58 (GenBank accession nos. AB516655, AB516656, and AB516657, respectively).

Southern blot analysis revealed that the clone pTm69 hybridized to a 6 kb
single major DNA fragment in the genomes of *T. monococcum* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD), but did not hybridize to genomic DNA of *Ae. tauschii* (DD) and *S. cereale* (RR) (Figure 1A). The single band of hybridization showed the presence of a tandem array that has equal 6 kb spacing in EcoO109I digests. The 6 kb tandem repetitive unit contained an unknown 225 bp sequence (accession no. AB516655). A search in blastn against all DNA databases of living organisms revealed that the repetitive element pTm69 only had partial similarities to several DNA segments in Bacterial Artificial Chromosomes (BACs) of mice, humans and zebrafish and did not share identity with class I or class II transposable elements or other known repetitive elements.

The clone pTm6 hybridized to several major DNA fragments in the genomes of *T. monococcum* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD), but did not hybridize to genomic DNA of *Ae. tauschii* (DD) and *S. cereale* (RR) (Figure 1B). The hybridization signal of pTm6 to the *T. monococcum* genomic DNA was very weak. The sizes of the major fragments were 4.6 kb in *T. monococcum* (AA), 2.4 kb, 4.1 kb and 4.6 kb in *T. turgidum* (AABB), and 2.1 kb, 4.2 kb and 5.0 kb in *T. aestivum* (AABBDD), indicating considerably restriction fragment length polymorphism among these species. The clone pTm6 contained a 265 bp sequence (accession no. AB516656) and showed highly identity (87%) to the integrase domain of barley Ty1-copia retrotransposon BARE-1 (Suoniemi et al. 1996). A blastn search using this repetitive element against all DNA databases identified segments in BACs from *T. monococcum* (AA), *T. urartu* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD), which contain the A genome.

Figure 2 shows a phylogenetic tree of the integrase domains found in the BACs and pTm6. The tree demonstrates high conservation of the integrase domain in A-genome-containing *Triticum* species and its large divergence in *Hordeum* species.

The other six clones (pTm58, pTm60, pTm66, pTm68, pTm71 and pTm85) showed strongly smeared hybridization signals in the EcoO109I digest on *T. monococcum* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD), but did not
hybridize to genomic DNA of *Ag. tauschii* (DD) and *S. cereale* (RR). A representative hybridization by the clone pTm58 is shown in Figure 1C. These *T. monococcum* sequences can be considered to be derived from a repetitive sequence family that was dispersed in the A genome against D and R genomes. Among the smeared signals, several sizes of distinct fragments were observed. The clone pTm58 contained a 217 bp sequence (accession no. AB516657) and showed 86% similarity to a region of the *Hordeum vulgare* Ty3-gypsy-like retrotransposon Romani-Hv (Suoniemi et al. 1998, Figure 3A). This sequence was found as several segments in BACs of *T. monococcum* (AA), *T. urartu* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD). However, the Romani-Hv designation was based on only two fragments, both of which were located in the same BAC, and the two fragments were presumed to be a single pair of long terminal repeats (LTRs) of the retrotransposon. The pTm58 family had a tRNA structure (Figure 3B), in which several sequences were conserved (for example, in the RNA polymerase III promoter, GUUCA in the TψC arm, and CCA at the 3’ end). Since tRNA structure has frequently been observed in retrotransposons such as SINE or LINE elements (Okada 1991; Murata et al. 1993; Ohshima et al. 1996), the pTm58 family may represent a retrotransposon-like sequence. The dispersed distribution throughout the genome shown by the smear hybridization on each genome blot corresponded well to its homology with retrotransposons (Figure 3A).

**Recovery of D-genome-specific repetitive sequence family by a subtraction method**

We established a DNA library in which the sequences common to wheat were subtracted (deleted) from the *Ae. tauschii* genome. Genomic DNA of *Ae. tauschii* randomly cleaved by sonication was mixed in excess with *MboI*-digested products of *T. turgidum*. The double stranded nucleotides were denatured at a high temperature into single strands, and then re-annealed as the double stranded DNA
in a phenol emulsion re-association mixture to recover rye-specific DNA elements restored with cohesive terminals.

The $Mbo$I fragments of *Ae. tauschii* (5 μg) were mixed with 15.0 μg of sonicated fragments of *T. turgidum*, annealed and desalted, then collected in 5.0-μg lots. From the collected DNA, 120 ng was used for ligation with 360 ng of pUC19 and yielded 253 recombinant plasmids. The $Mbo$I fragments (*T. turgidum*) and the sonicated fragments (*Ae. tauschii*) were mixed at a ratio of 1:3. Thus, 30.0 ng of the $Mbo$I fragments of *Ae. tauschii* should have been contained in the ligation solution. This means that 253 recombinant clones resulted from 30.0 ng of $Mbo$I fragments of *Ae. tauschii*; that is, 8,433 clones per μg. Ligation of 1 μg of the same $Mbo$I fragments into the *Bam*HI site of pUC19 by the shotgun method without the above-mentioned subtraction preparation yielded $3.28 \times 10^4$ recombinant clones. The number of clones generated after the re-annealing corresponds to 6.9% of those generated using the shotgun method. Thus, it appears that 93.1% of the $Mbo$I fragments randomly annealed with the sonicated fragments, whereas 6.9% of the $Mbo$I fragments were restored and able to ligate with the vector. The percentage of the clones obtained using the shotgun method was henceforth considered the restoration rate of the $Mbo$I fragments. Dot hybridization with the total DNA of *Ae. tauschii* or *T. turgidum* showed 5 of 253 clones with hybridization signals specific for *Ae. tauschii* that were stronger than control signals corresponding to 1,000 copies in the *Ae. tauschii* genome; these seemed to be the *Ae. tauschii* genome-specific repetitive sequences.

The core units of the screened repetitive clone were analyzed by Southern blot hybridization. Southern analysis indicated the hybridization specificity of the clone pAt1 for *Ae. tauschii* (2.6 kb and 4.4 kb fragments) against *T. turgidum* and rye (Figure 4). These RFLP (Restriction Fragment Length Polymorphism) fragments unique to the D genome could be used as a molecular tag for D genome chromatin. The nucleotide sequences were determined and clone pAt1 was found to have an unknown 86 bp sequence (accession no. AB516658) that did not share any identity with known class I or class II transposable elements or other known
repetitive elements. A search in blastn against all DNA databases of living organisms revealed that the repetitive elements had similarities to four unknown dispersed segments in BACs of wheat (e.g., AM932685.1 *Triticum aestivum* BAC No. TA3B95F5; and AM932684.1 *Triticum aestivum* BAC No. TA3B95C9). The 86 bp sequence was repeated twice in BAC No. TA3B54F7. The four wheat elements in the BACs consisted of the full structure of the repetitive element harboring the 86 bp segment: one in AM932685.1 that was 1,413 bp long, one in AM932680.1 that was 1,415 bp long, one in AM932680.1 that was 1,395 bp long, and one in AM932684.1 that was 1,850 bp long, showing 66% to 77% identity to each other. The former three core sequences in these BACs showed 95% identity (Figure 5). The regions flanking the repetitive elements showed no homology between the four BAC sequences. The entire structures of the elements did not share identity with class I or class II transposable elements or other known repetitive elements. The unknown new sequence did not contain terminal inverted repeats on its ends.

**Discussion**

Several repetitive DNA families have been cloned from methylated relic DNA since initial success by Appels and coworkers (1981, 1982, 1986). However, *Triticum* species contain repetitive sequences that combine units of different repetitive sequences irrespective of their methylation level (Flavell and Smith 1976; Smith and Flavell 1977; Bedbrook et al. 1980; Flavell et al. 1981; McIntyre et al. 1988). In our previous study, we sought to clone repetitive elements from nonmethylated regions by genomic subtraction; the sizes of the repetitive core units in the four kinds of rye-specific clones obtained were distinguished by *Eco*O109I (Tomita et al. 2009a). It was pertinent that all clones hybridized strongly to *Eco*O109I repetitive units of different lengths, namely 380 bp, 960 bp, 5 kb, and 5.5 kb. Because *Eco*O109I sites are multi-recognition sequences, 5′-PuGGNCCPy-3′ was present at high frequency in various repeated families. In the present study, total genomic DNA from *T. monococcum* was cleaved with
EcoO109I and different repetitive elements were then successfully isolated and classified into three classes of repetitive elements: pTm69, pTm6, and pTm58. These clones hybridized to the genomic DNA of *T. monococcum* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD), but did not hybridize to genomic DNA of *Ag. tauschii* (DD) and *S. cereale* (RR). The threshold used for specificity in screening was 1,000 copies, such that we only targeted highly repetitive sequences with a copy number greater than 1,000. Therefore, for the three A genomic clones that did not hybridize to the D genome, this means there were more than 1,000 copies of difference for these clones between the A and D genomes. Repetitive sequences such as these that show considerable differences in copy numbers between genomes could be useful molecular tags specific for each genomic chromatin.

The enzyme EcoO109I can generate polymorphic DNA fragments of unpredictable length because it recognizes the ambiguous motif 5′-PuGGNCCPy-3′, eventually representing 16 different heptamer motifs. EcoO109I is insensitive to (cytosine-5)-methylation of 5′-CpG-3′ sites, because none of the possible 16 different EcoO109I recognition motifs overlaps with this (cytosine-5)-methyltransferase recognition motif. In this study a new repetitive DNA element (pTm69) was successfully discovered by using the multiple cutter EcoO109I among only 21 highly repetitive clones with more than 1,000 copies screened from a small library (273 clones). EcoO109I is not biased against repeats and therefore is potentially useful for studying the organization and differentiation of large amounts of repetitive sequences in the Triticeae genome.

As described, a library containing DNA fragments specific for the D genome was produced by subtracting *T. turgidum* genomic DNA from genomic DNA of *Ae. tauschii*. After ligation of the *Ae. tauschii* MboI fragments, 253 recombinant plasmids were obtained. Subtraction by the deletion enrichment scheme was effective because the re-association rates of MboI fragments were estimated to be 6.9%, compared with shotgun cloning. From the restoration rate of the enzyme-digested fragments, it would seem that annealing of the mixture at a ratio
of 1:3 (restricted: sonicated fragments) resulted in this success. The insertion size of the repetitive clone was around 100–200 bp. A highly repetitive clone pAt1 that hybridized strongly to D genome DNA was screened further by differential dot blot hybridization. The clone pAt1 contained an unknown 86 bp sequence and we further identified a novel repetitive sequence family ranging in size from 1395–1850 bp from BAC sequences based on similarity to pAt1. This new repetitive sequence family did not show any similarity to known repetitive elements in Triticeae.

At 16.7 Gb, the wheat haploid genome is the largest among the Triticeae genomes. Angiosperm genomes vary greatly in size, with some species having less than 50 Mb of DNA per haploid genome and others having more than 85,000 Mb (Bennett and Leitch 1995). Within the grasses (Poaceae) there is also great variability in genome size: 450 Mb for rice, 2,500 Mb for maize, 5,000 Mb for barley, and 16,000 Mb for hexaploid wheat (Arumuganathan and Earle 1991). The amount of repetitive DNA, which comprises 70% of the cereal genomes, accounts for most of the variation in genome size (Flavell et al. 1977; Barakat et al. 1997; SanMiguel et al. 1996; Vicente et al. 1999; Myers et al. 2001; Wicker et al. 2001). Estimations of repetitive sequence content via C_0t analysis has revealed that 75% of the wheat genome is made up of repetitive sequences (Flavell et al. 1977), 16% of which are wheat-specific (Rimpau et al. 1978). The structures of pTm6 and pAt1 were quite different from those of known repetitive elements.

Several types of the repetitive DNA elements have been identified and studied. The LTR retrotransposons have greatly increased copy numbers (Doolittle and Sapienza 1980; Orgel and Crick 1980) and are therefore thought to have contributed greatly to expansions in genomes (Suoniemi et al. 1996; SanMiguel et al. 1996; Tikhonov et al. 1999; Bennetzen 2000; Park and Kazazian 2000). Over a broad range of organisms, retrotransposon copy number appears to be correlated with genome size. The small genome of the yeast Saccharomyces cerevisiae (13 × 10^6 bp) contains 51 full-length retrotransposons (Kim et al. 1998). In the large genome of maize (2,500 Mb), some retrotransposons have copy numbers
exceeding $2 \times 10^4$ per haploid genome (SanMiguel et al. 1996; SanMiguel and Bennetzen 1998; Wicker et al. 2001; Ramakrishna et al. 2002). All of the retrotransposons sequenced in maize have been inserted within the last 6 million years, leading to a doubling of the size of the maize genome (SanMiguel et al. 1998). It possible that plant genomes may also have expanded due to class II transposable elements, such as MITE and CACTA elements (Feschotte and Wessler 2002; Langdon et al. 2003; Wicker et al. 2003; Zhang et al. 2004). The GenBank database contains numerous different wheat transposable elements belonging to various classes; including Copia elements, Gypsy elements, LINE elements, SINE elements, CACTA elements, and MITE elements. In this study, the clones pTm6 and pTm58 showed similarity with retrotransposons. On the other hand, pTm69 and pAt1 were clearly different from both retrotransposons and DNA-type transposable elements. Despite extensive characterization of the repetitive elements in the Triticeae genome, pTm69 and pAt1 showed no similarity to any known repetitive elements, including the 350-family (Bedbrook et al. 1980; Appels et al. 1981), the 120-family (Bedbrook et al. 1980; McIntyre et al. 1990), the 5.3H3 family (Appels et al. 1986; McIntyre et al. 1988), the R173 family (Guidet et al., 1991; Rogowsky et al. 1991, 1992), pAs1 family (Rayburn and Gill 1986), the pSC250 family (Vershinin et al. 1995, 1996), and the pAt2 family (Tomita and Misaki 2010). These elements are therefore novel members of a genomic element family, as are LTR retrotransposons, SINE, MITE, and CACTA elements, which have contributed to plant genome structure and evolution.

The clones pTm69, pTm6, pTm58 and pAt1 were enriched in the wheat genome but were rare in the rye genome. These novel elements isolated here might serve as useful molecular markers for wheat chromatin in triticale. Moreover, dispersed elements pTm6 and pTm58 can yield many polymorphisms by sequence-specific amplified polymorphism (SSAP) or transposon display (Waugh et al. 1997), as has been shown with Revolver (Tomita 2008; Tomita et al. 2009c) and the R173 family of rye, which was used for DNA marker development.
Materials and Methods

Plant materials

The plant materials used in this study were *Triticum monococcum* (*2n = 2x = 14, AA); *Aegilops tauschii* (*2n = 2x = 14, DD); *T. turgidum* (*2n = 4x = 28, AABB); wheat *T. aestivum* L. cultivar Chinese Spring (CS) (*2n = 6x = 44, AABBDD); rye *Secale cereale* L. inbred pure line IR27 (*2n = 2x = 14, RR), developed from cultivar Petkus at Tottori University.

Isolation of A-genome-specific repetitive elements using *EcoO109I*

Genomic DNA was extracted from fresh leaves by the cetyltrimethylammonium bromide (CTAB) method from *T. monococcum*. The total genomic DNA of *T. monococcum* was completely digested with *EcoO109I*, the recognition sites of which consist of 5′-PuGGNCCPy-3′ multi-sequences and are present at high frequency in repetitive sequence families (Tomita et al. 2009a). The cohesive ends of the *EcoO109I*-digested fragments were converted to blunt ends by treatment with Klenow fragments. The blunted DNA fragments were ligated with the blunt ends of pUC19 linearized by *SmaI* restriction digestion, and subsequently transformed into competent *Escherichia coli* JM109 using 0.1 M CaCl₂ and PEG 600.

Recombinant plasmids were isolated by the alkali SDS method or by a single-step method (He et al. 1989); 1-μl aliquots were spotted onto nylon membranes, which were baked at 80°C for 3 h. The membranes were prehybridized for 3 h in 5×SSC, 2% blocking reagent (Roche Diagnostics, Basel, Switzerland), 0.1% N-lauroylsarcosine, 0.02% SDS, and 1 μg/ml sheared pUC109 DNA. Dot blot hybridization was performed with the total DNA of *T. monococcum* (5 ng) as a probe. The probe was labeled by a random primer method using digoxigenin-11-dUTP (Roche Diagnostics, Basel, Switzerland). The membranes were probed with the genomic DNA for 14 h at 68°C; the membranes
were then washed twice in 2×SSC with 0.1% SDS at room temperature for 5 min
and twice in 0.1×SSC with 0.1% SDS at 68°C for 15 min. Probe binding
hybridization signals were detected by using an anti-digoxigenin antibody labeled
with alkaline phosphatase (Roche Diagnostics, Basel, Switzerland) and was
visualized on X-ray film by light generation using
3-(2'-spiroadamantyl)-4-methoxy-4-(3-phosphoryloxy)-phenyl-1,2-dioxetane
(AMPPD). Only clones with hybridization signals stronger than control signals
corresponding to 1,000 copies in the *T. monococcum* genome were selected from
the dot blotting.

**Construction of D-genome-specific DNA libraries by genomic subtraction**

Genomic DNA was extracted from fresh leaves by the cetyltrimethylammonium bromide (CTAB) method from *Ae. tauschii* and *T. turgidum*. Repetitive sequences from wheat species were intricately differentiated by combining units of different repetitive sequences (Flavell and Smith 1976; Smith and Flavell 1977; Bedbrook et al. 1980; Flavell et al. 1981; McIntyre et al. 1988). Thus, to anneal minimum repetitive-sequence units, the genomic DNA of *Ae. tauschii* was completely digested into 2-kb fragments or less using the restriction enzyme *Mbo*I, which recognizes four nucleotides.

*Mbo*I-digested genomic DNA of *Ae. tauschii* was separated on 0.8–1.0% agarose gels, and 0.5–2.0 kb fragments were collected by freezing in liquid nitrogen (Koenen 1989). The collected DNA was mixed with genomic DNA of *T. turgidum* randomly cleaved into 1-kb fragments or less by sonication, and the mixed DNA was denatured in boiled water for 10 min. To facilitate annealing, the mixed DNA was settled in a revolving flask with 4 ml of phenol emulsion buffer (8% phenol, 1.25 M NaClO₄ and 0.12 M Na₂HPO₄, pH 6.8), and the flask placed in a rotary evaporator for 72 h at room temperature. In the process of forming double stranded DNA, the *Ae. tauschii Mbo*I fragments with common sequences to *T. turgidum* DNA were associated with excessive amounts of wheat fragments of different lengths and terminal forms, whereas *Mbo*I fragments with repetitive
sequences specific for *Ae. tauschii* were re-annealed to restore the double strands with cohesive terminals. When sticky-end plasmids with complementary sequences to the restored DNA were prepared as vectors, only rye-specific double strands with cohesive ends were ligated. This procedure is designed to recover *Ae. Tauschii*-specific repetitive sequences, which will self-anneal to form cohesive ends adjusted to the vector ends. This solution of DNA reconstituted by annealing was passed through Sephadex G-25 to eliminate the phosphate salt. After extraction with phenol, the recovered DNA was used to ligate with pUC19 linearized with *Bam*HI. Competent *Escherichia coli* JM109 was then transformed using 0.1 M CaCl₂ and PEG 600.

Recombinant plasmids were isolated and an aliquot of 1.5 µl was spotted on two sheets of nylon membranes. Dot blot hybridization was performed using either the total DNA of *Ae. tauschii* (100 ng) or the total DNA of *T. turgidum* (200 ng) as a probe. Probe labeling, hybridization and detection were conducted according to the above section. One membrane was probed with genomic DNA of *Ae. tauschii* and the other was probed with *T. turgidum* DNA. Clones with strong hybridization to total genomic DNA of *Ae. tauschii* were selected from the differential dot blotting. The repetitive clones were scored as those with hybridization signals stronger than control signals corresponding to 1,000 copies in the *Ae. tauschii* genome.

**Southern analysis**
Genomic DNA (10 µg) of *T. monococcum, T. turgidum, Ae. tauschii, T. aestivum* var. CS, and *S. cereale IR27* was digested to completion with the restriction enzyme *EcoO109I* or *DraI*; subjected to electrophoresis on 1% agarose gels; and then transferred and cross-linked to nylon membranes. The membranes were then probed with the repetitive sequence clones. The membranes were prehybridized at 65°C for 3 h in hybridization solution containing 5×SSC, 2% blocking reagent (Roche Diagnostics, Basel, Switzerland), 0.1% N-lauroylsarcosine, and 0.02% SDS. Wheat repetitive probes (25 ng) labeled with digoxigenin-11-dUTP was
added to the prehybridization solution and the mixture was allowed to hybridize at 65°C for 14 h. Blots were washed twice in 2×SSC with 0.1% SDS at room temperature for 5 min and twice in 0.1×SSC with 0.1% SDS at 65°C for 15 min before exposure to X-OMAT film (Eastman Kodak, NY, USA) using CDP-Star (Roche Diagnostics, Basel, Switzerland). Stringency conditions were comparable to those of the dot blot analysis.

DNA sequencing and phylogenetic analysis

To determine the structure of the repetitive elements, which were isolated using EcoO109I or genomic subtraction, the repetitive sequence clones pTm6, pTm69, pTm58 and pAt1 were fully sequenced by PCR cycle reactions on both strands and using an ALF Express sequencer (GE Healthcare, Buckinghamshire, UK). Overlapping fragments were assembled with ALF Express assembler software. Completed nucleotide sequences were compared with sequences in the non-redundant GenBank+EMBL+DDBJ databases by using BLASTN homology search software (Altschul et al. 1997). The sequence alignment was determined using the computer program DNASIS Pro version 2.10 (Hitachi, Tokyo, Japan). A non-rooted phylogenetic tree was constructed using MEGA version 4.0.2 (Kumar et al. 2004). The neighbor-joining method (Saitou and Nei 1987) was conducted with Kimura’s 2-parameter distances (Kimura 1980).
References


Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acid Res.* **20**, 3639–3644.


Nagy ED, Lelley T (2003). Genetic and physical mapping of sequence-specific amplified polymorphic (SSAP) markers on the 1RS chromosome arm or rye in a


Tomita M, Shinohara K, Morimoto M (2008). *Revolver* is a new class of


Figure Legends

Figure 1. Southern blot analysis of the three classes of A-genome-specific repetitive sequences pTm69 (A), pTm6 (B) and pTm58 (C) and the non-genome-specific clone pTm5 (D).

Clones pTm69, pTm6 and pTm58 hybridized to genomic DNA of *T. monococcum* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD), but not *Ae. tauschii* (DD) or *S. cereale* (RR).

The hybridization signal of pTm6 on the *T. monococcum* genome was very weak.

The sharp main signal of pTm69 indicated that the repetitive sequences were derived from tandem repeat sequence family, while the smeared hybridization signals of pTm58 indicate that the repetitive sequence was derived from dispersed repetitive sequence family.

Figure 2. Neighbor-joining tree of pTm6 related sequences in the Triticeae. BAC accession numbers are indicated alongside species names; numbers on branches indicate the boot strap values.


Figure 3. Structure of the dispersed repetitive sequence family pTm58.

(A) Sequence alignment from pTm58 of *T. monococcum* and the Ty3-gypsy-like retrotransposon Romani of *H. vulgare* (AJ002618, nucleotides 178-394; AJ002619, nucleotides 177-393).

(B) tRNA structure of pTm58.

Figure 4. Southern blot analysis of *Dra*I digests probed with the clone pAt1. Two
RFLP fragments were specific to the genomic DNA of *Ae. tauschii* (DD) but not to *T. turgidum* (AABB) or *S. cereale* (RR).

Figure 5. Sequence alignment of a new dispersed repetitive sequence family in wheat BACs (AM932685, AM932680.1-1, AM932680.1-2) identified by clone pAt1.
Figure 1. Southern blot analysis of the three classes of A-genome-specific repetitive sequences pTm69 (A), pTm6 (B) and pTm58 (C) and the non-genome-specific clone pTm5 (D). Clones pTm69, pTm6 and pTm58 hybridized to genomic DNA of *T. monococcum* (AA), *T. turgidum* (AABB) and *T. aestivum* (AABBDD), but not *Ae. tauschii* (DD) or *S. cereale* (RR). The hybridization signal of pTm6 on the *T. monococcum* genome was very weak. The sharp main signal of pTm69 indicates that the repetitive sequences were derived from a tandem repeat sequence family, while the smeared hybridization signals of pTm58 indicate that the repetitive sequence was derived from dispersed repetitive sequence family.
Figure 1. continued.
Figure 2. Neighbor-joining tree of pTm6 related sequences in the Triticeae. BAC accession numbers are indicated alongside species names; numbers on branches indicate the boot strap values.

Figure 3. Structure of the dispersed repetitive sequence family pTm58.

(A) Sequence alignment of pTm58 from *T. monococcum* and the Ty3-gypsy-like retrotransposon *Romani* of *H. vulgare* (AJ002618, nucleotides 178–394; AJ002619, nucleotides 177–393).

(B) tRNA structure of pTm58.
Figure 4. Southern blot analysis of *DraI* digests probed with the clone pAt1. Two RFLP fragments were specific to the genomic DNA of *Ae. tauschii* (DD) but not to *T. turgidum* (AABB) or *S. cereale* (RR).
Figure 5. Sequence alignment of a new dispersed repetitive sequence family in wheat BACs (AM932685, AM932680.1-1, AM932680.1-2) identified by clone pAt1.
Fig. 5. Continued.
Fig. 5. Continued.
Fig. 5. Continued.
Fig. 5. Continued.