A Centromeric Tandem Repeat Family Originating From a Part of Ty3/*gypsy*-Retroelement in Wheat and Its Relatives

Zhi-Jun Cheng* and Minoru Murata*,†,[1]

**Core Research for Evolutionary Science and Technology Program, Japan Science Technology Corporation, Tokyo 103-0027, Japan and* †*Research Institute for Bioresources, Okayama University, Kurashiki 710-0046, Japan*

Manuscript received October 4, 2002
Accepted for publication February 17, 2003

ABSTRACT

From a wild diploid species that is a relative of wheat, *Aegilops speltoides*, a 301-bp repeat containing 16 copies of a CAA microsatellite was isolated. Southern blot and fluorescence *in situ* hybridization revealed that ~250 bp of the sequence is tandemly arrayed at the centromere regions of A- and B-genome chromosomes of common wheat and rye chromosomes. Although the DNA sequence of this 250-bp repeat showed no notable homology in the databases, the flanking or intervening sequences between the repeats showed high homologies (>82%) to two separate sequences of the *gag* gene and its upstream region in *cereba*, a Ty3/*gypsy*-like retroelement of *Hordeum vulgare*. Since the amino acid sequence deduced from the 250 bp with seven CAAs showed some similarity (~53%) to that of the *gag* gene, we concluded that the 250-bp repeats had also originated from the *cereba*-like retroelements in diploid wheat such as *Ae. speltoides* and had formed tandem arrays, whereas the 300-bp repeats were dispersed as a part of *cereba*-like retroelements. This suggests that some tandem repeats localized at the centromeric regions of cereals and other plant species originated from parts of retrotransposons.

CENTROMERES play an essential role in precise segregation of sister chromatids at mitosis and meiosis. The DNA structure and associated proteins have been extensively studied in yeasts and humans (reviewed by Pidoux and Allshire 2000). Data on centromeres have been accumulating in higher plants; however, the centromeric DNA constitution is still unclear (reviewed by Murata 2002). In *Arabidopsis thaliana*, centromeric regions are composed mainly of tandem repeats, called pAL1 or the 180-bp family (Murata *et al.* 1994; Brandes *et al.* 1997; Arabidopsis Genome Initiative 2000), but retrotransposons and middle repetitive sequences have also been identified (Brandes *et al.* 1997; Copenhaver *et al.* 1999). In cereals, two repetitive DNA sequences, CCS1 (Aragón-Alcaide *et al.* 1996) and pSau3A9 (Jiang *et al.* 1996), were first found to localize preferentially at the centromeric regions and were shown later to have similarities to parts of Ty3/*gypsy*-type retrotransposons (Miller *et al.* 1998; Presting *et al.* 1998). Large-scale sequencing analyses of cosmid, bacterial artificial chromosome (BAC), and/or yeast artificial chromosome (YAC) clones revealed that multiple copies of partial or whole Ty3/*gypsy*-type retrotranposons are dispersed in the centromeric regions of barley (Hudakova *et al.* 2001), wheat (Fukui *et al.* 2001), and rice (Nonomura and Kurata 2001). Meanwhile, tandem repeat sequences with <1 kb unit have also been found in the centromeric DNA of common wheat (Kishii *et al.* 2001), maize (Ananiev *et al.* 1998), pearl millet (Kamm *et al.* 1994), rice (Dong *et al.* 1998), and sugar cane (Nagaki *et al.* 1998). Contrary to the centromeric retrotransposons, the DNA sequences of which are conserved, these tandem repeats are not conserved even within the same species (Ananiev *et al.* 1998; Chen *et al.* 2000; Brent *et al.* 2001; Kishii *et al.* 2001). However, it was recently shown that 65 kb to 2 Mb clusters of a tandem repeat family (Cent0) with a 155-bp unit are located within the function domains of all rice centromeres (Cheng *et al.* 2002). This suggests that centromeric tandem repeats confer centromere functions in cereals as indicated in *A. thaliana* and other species.

In this work, we identified a novel tandem repeat family with a unit size of ~250 bp in *Aegilops speltoides* and aimed to characterize its hybridization in the centromeric regions of A- and B-genome chromosomes of wheat and rye chromosomes as well as those of *Ae. speltoides* chromosomes. The origin and amplification mechanisms of this centromeric repeat family are also discussed.

MATERIALS AND METHODS

Plant materials: Two lines of *Ae. speltoides* var. *typica* (accession nos. KU5727 and KT115-1), kindly provided by Drs. T. R. Endo (Kyoto University) and K. Tsunewaki (Fukui Prefectural University), respectively, were used as genomic *in situ* hybridization (GISH) probes and a source of repetitive DNA sequences. In addition, the following materials were also used: the octoploid triticale line Y4683 (Cheng and Murata 2002), *Secale cereale* "Petkus," *Triticum monococcum* var. *vulgare* (KT3-1), *Ae. squarrosa* var. *strangulata* (KT120-5), *T. durum* var. *Rei-*

[1]*Corresponding author:* Research Institute for Bioresources, Okayama University, Kurashiki 710-0046, Japan.
E-mail: mmura@rib.okayama-u.ac.jp

chenvachii, *T. aestivum* "Chinese Spring" (CS), *Oryza sativa* "Nipponbare," *Hordeum vulgare* "Ishukushirazu," and *Avena sativa* "Albion."

DNA extraction, cloning, and sequencing: All genomic DNA were extracted from 1- to 2-week-old seedlings by the Nucleon Phytopure DNA extraction kit (Amersham Bioscience, Arlington Heights, IL). One microgram of *Ae. speltoides* (KT115-1) DNA was partially digested with 0.02 unit of restriction enzyme *Sau*3AI at 37° to obtain fragments of ~2 kb in length, ligated to pBluescript II SK(+) (Stratagene, La Jolla, CA) linearized with *Bam*HI, and transformed to *Escherichia coli* XL10-Gold (Stratagene). Colonies were transferred to nylon membranes (Hybond N^+; Amersham Bioscience) and hybridized with digoxigenin (DIG)-labeled genomic DNA of CS to identify repetitive DNA sequences. The selected clones were sequenced and investigated by fluorescence *in situ* hybridization (FISH) to find their chromosomal locations.

PCR amplification and sequence data analysis: The centromeric repeat sequences were obtained by PCR performed with various kinds of genomic DNA as a template, under the conditions of 94° for 30 sec, 59° for 30 sec, and 72° for 30 sec, for 30 cycles. Then the product was purified with SUPREC PCR (Takara Biomedical, Berkeley, CA), ligated to pGEM T-easy vector (Promega, Madison, WI), transformed to *E. coli* XL10-Gold, and sequenced. The primers used were ZCF1 (5′-CTG GCCTTGAGAAGACGTTC-3′), ZCF3 (5′-CGTTCGAAACAAA GCTCGAT-3′), ZCF4 (5′-GATTATGCGGGAGATTACGAGG-3′), ZCF7 (5′-TCAGCAAATGATGTCTGACCA-3′), ZCR1 (5′-CCTCGTAATCTCCCGCATAA-3′), ZCR2 (5′-GAACGTCTT CTCAAGGCCAG-3′), and ZCR7 (5′-GATGGTCATAATCCC GTACCTG-3′). All sequence data were analyzed by SeqEd version 1.0.3 (ABI, Columbia, MD) and GENETYX-MAC-ATSQ-3.1 (Software Development) software for data processing and GENETYX-MAC 10.1 (Software Development) for homology comparison. We searched the nonredundant nucleic acid sequence database of NR-NT with BLASTN and BLASTX for all sequences. The nucleotide sequences reported here have been registered at DDBJ/EMBL/GenBank under accession nos. AB088401–AB088402 and AB099945–AB099949.

Genomic Southern blot hybridization: Three micrograms of genomic DNA was digested, electrophoresed and transferred to nylon membranes (Hybond N; Amersham Bioscience), hybridized in hybridization buffer (DIG Easy Hyb; Roche, Indianapolis) with DIG-labeled probes (Roche), and detected by chemical luminescent signals according to the manufacturer's instructions.

Chromosome preparation and FISH: Germinating seeds were placed at 4° for 24 hr to synchronize cell divisions and transferred to room temperature (RT; 25°). After 24 hr, root tips were collected at the root length of 1.5–2.0 cm and pretreated in ice-cold water for 24 hr. Then they were fixed in ethanol-acetic acid (3:1) and stored at −20°. Chromosome preparations for FISH and GISH were made according to the air-drying technique by Murata *et al.* (1992).

The clones isolated in this study and clone pSc74 (Bedbrook *et al.* 1980), genomic DNAs of rye, *T. monococcum*, *Ae. squarrosa*, and CS were labeled with biotin-14-dUTP (GIBCO BRL, Gaithersburg, MD) or DIG-11-dUTP (Roche) by nick translation. FISH and GISH were carried out using the procedure of Murata *et al.* (1992, 1997) with minor modifications. In brief, 20 μl of hybridization mixture consisting of 50% (v/v) deionized formamide, 10% (v/v) dextran sulfate, 2× SSC (0.3 M NaCl, 0.03 M sodium citrate), and one or two labeled probes (~40 ng each) was used. Avidin-FITC (Roche), streptavidin-Cy3 (Sigma, St. Louis), anti-DIG-fluoroscein (Roche), and/or anti-DIG-rhodamine (Roche) were applied to detect hybridized probes. Fluorescent images were obtained using a fluorescence microscope (Axioskop; Carl Zeiss, Thornwood, NY) with UV-, B-, and triple-band pass filters (Carl Zeiss filter nos. 01, 09, and 25, respectively) and a color-chilled CCD camera (C5810; Hamamatsu Photonics, Bridgewater, NJ). The images were processed using Photoshop 6.0 (Adobe).

RESULTS

Identification and characterization of centromeric repeats from *Ae. speltoides*: Genomic DNAs isolated from two lines of *Ae. speltoides* were labeled with DIG and hybridized together with biotin-labeled rye-specific pSc74 to mitotic metaphase chromosomes of octoploid triticale Y4683. One of the genomic DNA probes (KT115-1) showed a clear banding pattern (Figure 1A). Fourteen chromosomes of common wheat, identified as B-genome chromosomes, showed the strongest hybridization, and distinct signals appeared at the centromeric regions of rye and some other wheat chromosomes. The rye chromosomes were identified by the green telomeric signals from pSc74 probe.

To isolate the possible repetitive DNA sequences that hybridized preferentially to the centromeric regions, we constructed a *Sau*3AI partial-digested genomic DNA library from *Ae. speltoides* KT115-1 and screened it with a DIG-labeled CS genomic DNA probe. After stringent washing (0.1× SSC, 68°), 18 out of 500 clones showed relatively strong signals. All 18 clones were checked subsequently by FISH, and only clone 307-5 showed a FISH pattern similar to the previous GISH pattern (data not shown). The insert was 1083 bp in length and was divided into two parts: 782 bp and 301 bp by a *Sau*3AI site. Sixteen copies of CAA microsatellite were found in the latter part (Figure 2).

Genomic Southern hybridization with clone 307-5 as a probe showed a ladder pattern with an ~250-bp repeat unit in the total DNA of *Ae. speltoides* KT115-1 digested with *Sau*3AI (data not shown). This suggested that only a part of the insert is arrayed in tandem. To find the tandem sequence, the clone 307-5 DNA was digested with *Sac*I, which cut once at the 684-bp site (Figure 2), self ligated, and transformed. FISH with the resultant clone pBs301 as a probe, which contained a 301-bp *Sau*3AI fragment from the 3′-end of 307-5, showed a pattern similar to that of 307-5 (data not shown). pBs301 was also used as a probe in Southern hybridization to *Sau*3AI-digested DNA of CS, durum wheat, and five other diploid species (Figure 3). All species except *Ae. squarrosa* showed ladder patterns with ~250-bp repeat units. *S. cereale* had slightly faint 250-bp bands, whereas *T. monococcum* had a strong 500-bp band corresponding to a 250-bp-unit dimer. Both lines of *Ae. speltoides*, especially KU5725, showed a heavy smear, but the ladder patterns could be observed when the exposure time was short.

The smear observed in *Ae. speltoides* lanes was probably produced by the microsatellites of $(CAA)_{16}$ in the pBs301 probe, since some microsatellites are known

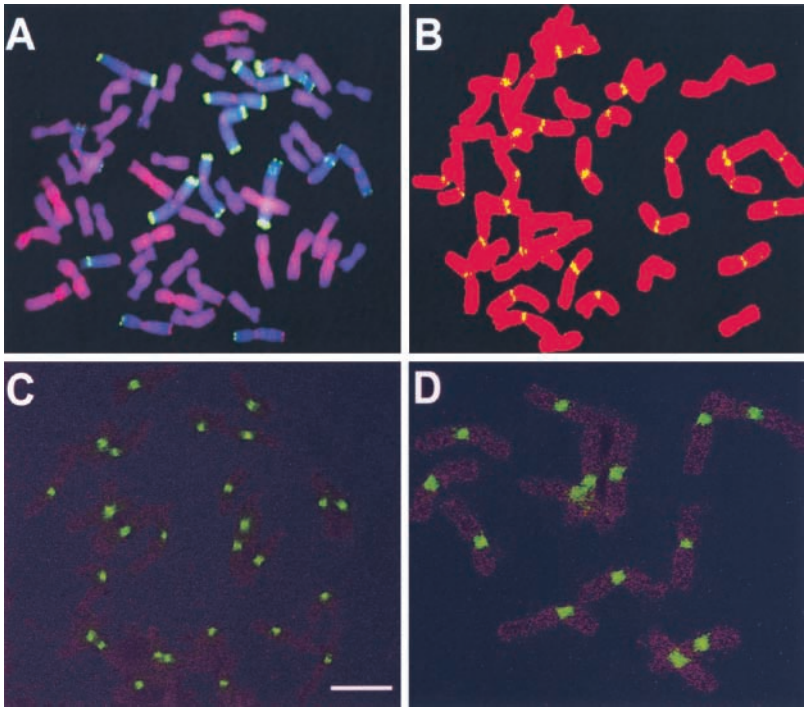

FIGURE 1.—GISH and FISH analysis showing localization of the repetitive DNA sequences from *Ae. speltoides*. (A) Mitotic chromosomes of octoploid triticale (Y4683) hybridized with biotinylated *Ae. speltoides* genomic DNA and DIG-labeled pSc74, detected with streptavidin-Cy3 (red) and anti-digoxigenin-fluorescein (green), respectively. (B–D) Mitotic chromosomes of common wheat CS counterstained with propidium iodide (B), durum wheat (C), and rye (D), respectively. All were hybridized with DIG-labeled CS2 clone and detected by anti-DIG-fluorescein. Bar, 10 μm.

to cause smearing in Southern blotting and dispersed signals in FISH (SCHMIDT and HESLOP-HARRISON 1996). This assumption was also supported by the size of a *Sau*3AI fragment in pBs301, which is ~250 bp when the microsatellite repeats (48 bp) are subtracted. To identify the putative 250-bp repeat and remove the effect of the microsatellite, we designed PCR primers of ZCF1 and ZCR1 to amplify the 288-bp fragment (named pBs301-1) of pBs301 (Figure 2). As a result, two bands of ~250 and 300 bp in length were amplified in almost all species used (Figure 4A).

To investigate the copy numbers of the microsatellite, we cloned and sequenced some of the PCR products. Two clones from CS (CS2, 245 bp and CS3, 253 bp; Figure 5) and two clones from *Ae. speltoides* KT115-1 (253 and 247 bp) contained only two to four copies of CAA microsatellite. All of these clones hybridized preferentially to the centromeric regions of 28 chromo-

```
GATCAAAAAG AGAGAAGAAG CCATCTAGCT AATAACTATG GACCTGTCGT GGAATTAACA   60
CGTGAGATGT CATAGTGTGA GGACTTAGTC GTGAGGCCAA CACATCTATG TGGTAGCTTG  120
AGAGGGGTTG AGCGGAATCG AGAAACACAA CACAAGTACG AGGATTTATG ACAGCTTTGG  180
GCCCCGGGAA ACATCATCCG GTAATAGCCC TACATGCTGT TTGTGGCTAG GTCTCATTAT  240
CATCACGAGT GAGTCGCCGT AAACCGGCTC TCCTCTAGTT GTGTCTAGCC CTAGAGATTG  300
TTTCTTGTTG CTTGTCCCTC TTTGGGGAGC CCTGCCCCTC CTTATATAAG TTGAAGGGGC  360
GGGTTACATG TAGAGTCCAT CTCGGACTAA GACTTAAACT ATTTTGACTT CCTTTCCTGG  420
GCTTCTTAAC GTCTGCCGGT TTAACATTGT TTGTCGGTTT ACCGTCTGTC TTAGCCATCT  480
GTCTTGACTC TCTGTCTTAA CTGTCCGCCG GTTTACCATC CGCCGGTTTA CCAAGTCCAA  540
GCCGGTTTAC AATTCCAGCC GGGTCATACC GCGGGGTATA TCCCCGACAT TAGCCCCCAG  600
TTTAATTTGA ATTTATCCAT TTTAAACTCG AAGAACAGTT TATACATCTC CACGGGCGTC  660
GCACCCATGA AGTGTTGGAG AGCTCGCGGT CCCTCCGGAT TGAATTCCCA CAGGCGGAGG  720
GGGCGACGTT TGCAAGGCAG AAGGCGTCGA ATCAGCATGA CCTGCGCCAC TACAACCAGA  780
TTGATCACTG GCCTTGAGAA GACGTTCGAA ACAAAGCTCG ATAACAGATT TAATGAACTG  840
CTTGCGCGTC TTCCACCACC GGTTGCACCT GCCGCACCTC TGCAACAACA ACAACAACAA  900
CAACAACAAC AACAACAACA ACAACAACAA TGACTACCTC CACATCGCGA AACAACCCTC  960
CACCGAGCGA GCCGTGTTCC TCTTCAACCT GGCCAAACTG TTGGTGCTGC TGTTGATAAT 1020
TCTGTGGCTC CTGCTGCTGA TGGGGAGGAG GATGATTATG CGGGAGATTA CGAGGATGAG 1080
GTTGATC                                                           1087
```

FIGURE 2.—The sequence of clone 307-5 (DDBJ/EMBL/GenBank accession no. AB088401). Positions of *Sau*3AI and *Sac*I sites are boxed. A pair of primers (ZCF1 and ZCR1) and a microsatellite array $(CAA)_{16}$ are arrowed and underlined, respectively.

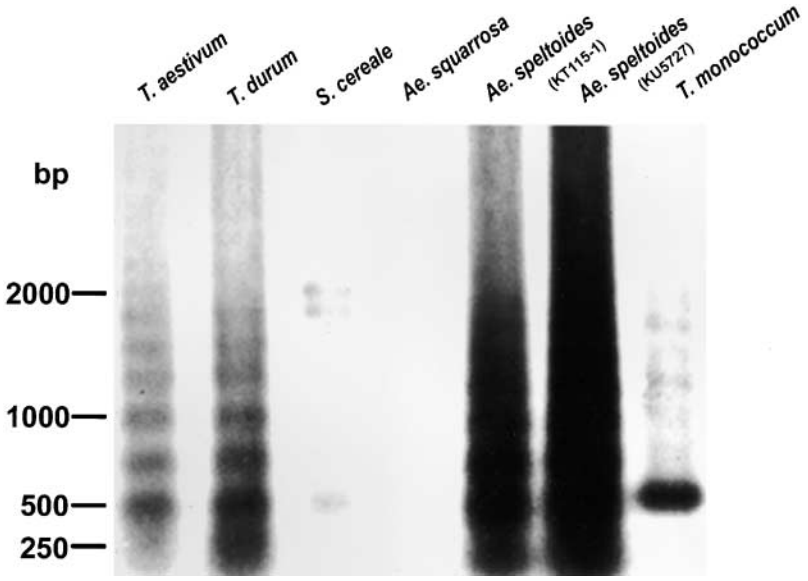

FIGURE 3.—Southern blot hybridizations of *Sau*3AI-digested genomic DNA of nine cereal species with pBs301 as a probe.

somes in common wheat (Figure 1B), 28 in durum wheat (Figure 1C), and 14 in *Ae. speltoides* (data not shown) or in rye (Figure 1D). In Southern blot hybridization to *Sau*3AI-digested genomic DNA, both 250-bp monomeric and 500-bp dimeric bands appeared without smearing in common wheat, durum, and *Ae. speltoides*, and a 500-bp band appeared in *T. monococcum* (data not shown). Although the lengths of monomers were variable (245–253 bp), the sequences appeared to be tandemly arranged within the centromeric regions in A- and B-genome chromosomes of wheat and R-genome chromosomes of rye. By contrast, no signals were found in rice, barley, millet, or oats (data not shown).

As shown in a sequence alignment of five PCR clones of CS to pBs301-1 of *Ae. speltoides* (Figure 5), all clones contained more than one copy of a CAA microsatellite

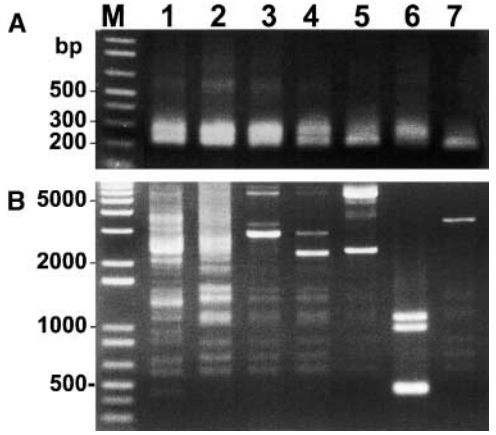

FIGURE 4.—Agarose electrophoresis of PCR products amplified by the primer combinations of ZCF1 + ZCR1 (A) and ZCR2 + ZCF3 (B). Genomic DNAs used as a template were: (1) CS (*T. aestivum*), (2) *T. durum*, (3) *Ae. speltoides* (KT115-1), (4) *Ae. speltoides* (KU5727), (5) *T. monococcum*, (6) *Ae. squarrosa*, and (7) *S. cereale*.

and its derivatives. The most common substitutions were a transversion from A to T, but A → G and C → T transitions also occurred. Compared with the variation of the microsatellite, the sequences flanking the microsatellites were conserved, although a 35-bp deletion in CS15, a 9-bp deletion in CS2, and short 1- to 4-bp deletions were in all clones. Similarly, in five PCR clones from *Ae. speltoides* (KT115-1), 3–16 copies of the CAA microsatellite were found. However, ~20 nucleotides just upstream the microsatellite arrays were less conserved than those of CS clones. These and other data indicated that ~250-bp sequences having a low copy number of CAA microsatellites are arrayed in tandem, but ~300-bp sequences, which have more copies, are located separately.

Intervening sequences among the 250- and 300-bp repeats: To amplify intervening sequences between the 250- and 300-bp repeats, we designed three different PCR primers: ZCF3, ZCF4, and ZCR2 from the conserved sequences among CS and pBs301-1 clones (Figures 5 and 6A). The combination of ZCF3 and ZCR2 amplified >10 bands ~450 bp to 5 kb in length from CS DNA (Figure 4B, lane 1), while the primer ZCF4 and ZCR2 combination produced only 2 faint bands. Similar amplification results were also obtained using other genomic DNA as a template (Figure 4B, lanes 2–7). The products from *Ae. speltoides* (KT115-1), *Ae. Squarrosa*, and CS were cloned and end sequenced. Out of 19 clones, 16 contained sequences homologous to parts of *cereba*, a Ty3/*gypsy*-like retroelement of *Hordeum vulgare* (HUDAKOVA *et al.* 2001; GenBank accession nos. AY040832 and AF078801; Figure 6A). Two separate regions of the *cereba* DNA showed high homologies (>82%) to the PCR products amplified from a primer combination of ZCF3 and ZCR2: one is an 80-bp 5′-region of *gag* gene (nucleotide position 1399–1479 in AY040832) and its upstream noncoding regions [5′-

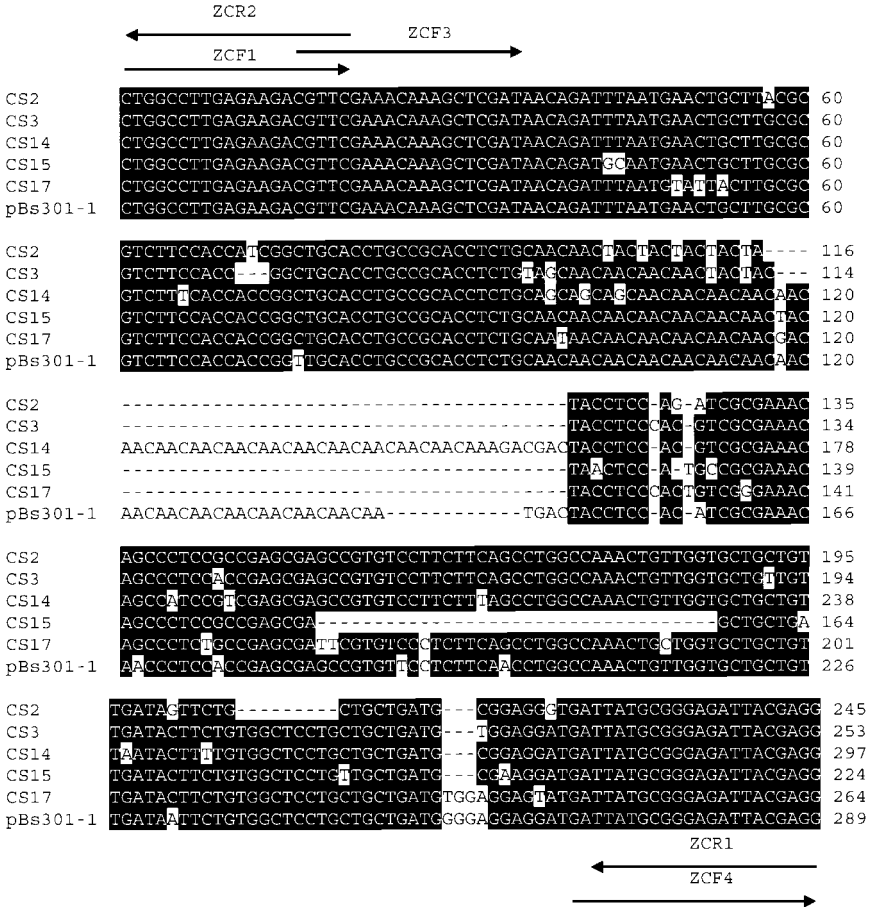

FIGURE 5.—DNA sequence alignment of five clones from common wheat and pBs301-1. The conserved nucleotides are in black boxes. The PCR primers used are also indicated.

untranslated region (UTR)] in different lengths; another is the coding regions 34–105 bp long in the *gag* gene, all of which started at position 1926 (Figure 6A). This revealed a distinct nonhomologous sequence ~450 bp long between the two homologous regions in the *gag* gene. Although no homology at the DNA sequence level was found between the sequence 1479–1926 of *cereba* and the 250- to 300-bp repeat such as pBs301, the amino acid sequence deduced from pBs301 showed ~53% similarity (41% identity) to that from the DNA sequence 1531–1824 of *cereba* (Figure 6B). This suggests that the 250- and 300-bp repeats had originated from Ty3/*gypsy*-like retrotransposons like *cereba* in *H. vulgare* and that there are *cereba*-like retroelements are in the A and B genomes of wheat, but their sequences corresponding to 1531–1824 in *cereba* (AY040832) vary much from that of *cereba*. To confirm this finding, we designed a new pair of PCR primers on the basis of the *cereba* sequences (GenBank accession no. AF078801), ZCF7 and ZCR7, which are expected to amplify a 458-bp fragment (3292–3749 in *cereba*, no. AF078801) or a 485-bp fragment (1467–1951 in *cereba*, no. AY040832; Figure 6A). The primer pair could amplify the expected size fragments from barley, having almost the same sequence as *cereba*, while all four PCR products from *Ae. speltoides* (KT115-1) contained the 250- to 300-bp sequences. The DNA sequence of clone ZC7-1 was almost identical to that of pBs301-1 except for the copy number of CAA microsatellite, which was reduced to five.

In the PCR products from CS as well as *Ae. speltoides*, however, we found highly conserved ~40- and 48-bp sequences at the junctions of the 250- to 300-bp repeats and *cereba*-like sequences, showing no homology to that of either *cereba* or pBs301 (Figure 6C). The origin of the two sequences is uncertain, since no sequences showing overall homology to either 40 or 48 bp were found in the DNA databases.

DISCUSSION

Ae. speltoides ($2n = 2x = 14$, genome constitution SS) has been shown to have the highest genetic affinity to the B genome of common wheat (reviewed by DVORAK and ZHANG 1990; FRIEBE and GILL 1996; TSUNEWAKI 1996). In our study, genomic DNA from *Ae. speltoides* painted B-genome chromosomes more heavily than A-genome chromosomes and produced interstitial and pericentric bands (Figure 1A). This GISH pattern was produced mainly by a 250-bp centromeric repeat and a CAA microsatellite, since the GISH pattern was very similar to the FISH pattern obtained with the pBs301 clone containing both the 250-bp repeat unit and 16

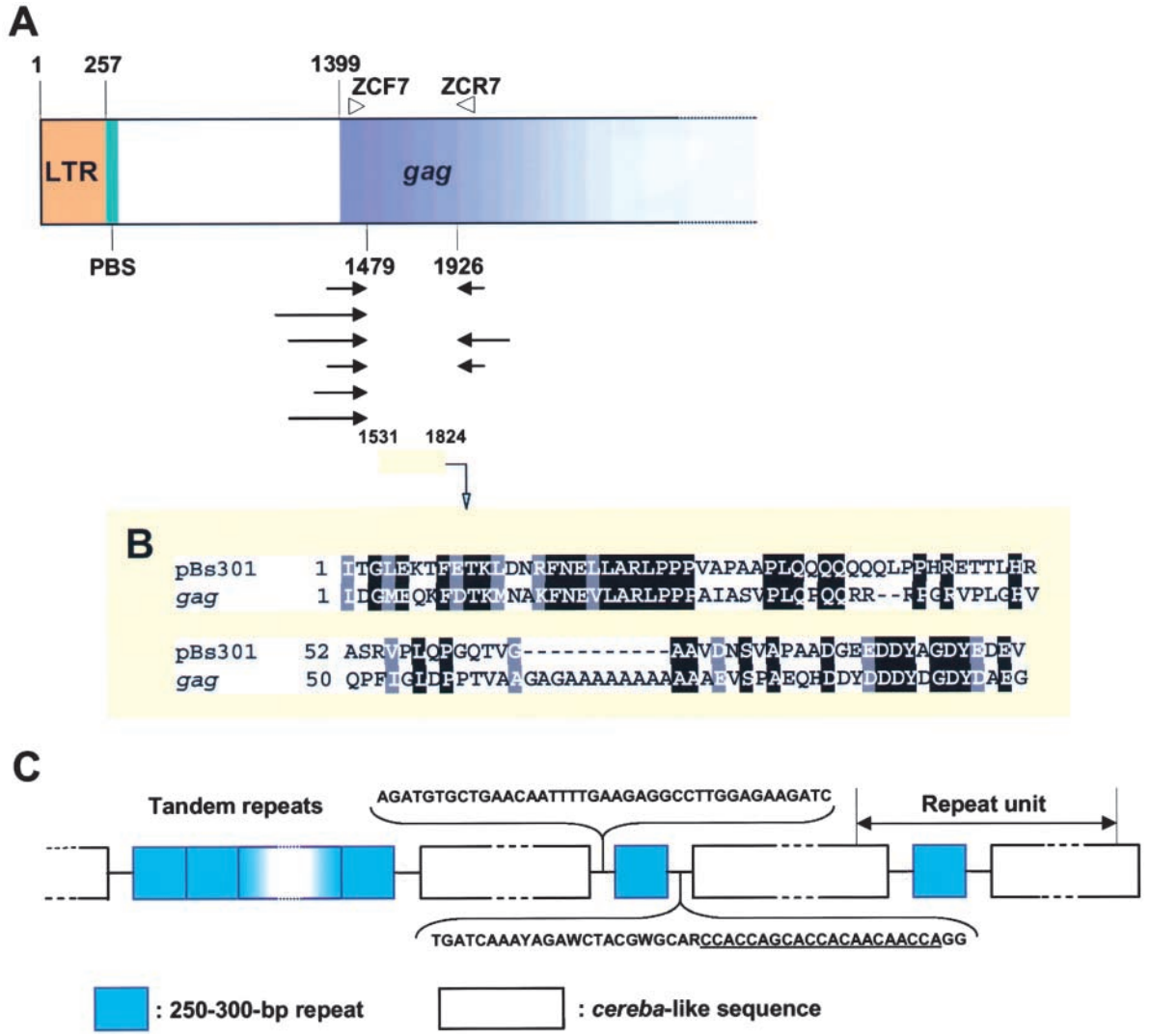

FIGURE 6.—Schematic representations showing the relationship between the 250- and 300-bp repeats and *cereba*, a Ty3/*gypsy*-like retroelement of *H. vulgare*. (A) A part of *cereba* containing a long terminal repeat (LTR), a primer-binding site (PBS), and a *gag* gene. Numbers correspond to the nucleotide positions of a *cereba* (accession no. AY040832). Arrows and arrowheads indicate the regions having high homologies (>82%) to the PCR products amplified by a combination of ZCF3 and ZCR2 primers. (B) An alignment of amino acid sequences deduced from pBs301 and a part of the *gag* gene (1531–1824 of AY040832). (C) Estimated organization of the 250- and 300-bp repeats and nonhomologous sequences found in the up- and downstream regions of the 250- to 300-bp repeats.

copies of CAA microsatellite as a probe. Furthermore, the 250-bp repeat with a few copies of the microsatellite such as pBs301-1 hybridized only to centromeric regions of A-, B-, and R-genome chromosomes as well as S-genome chromosomes (Figure 1).

Several tandem repeat families have been reported to be present in the centromeric regions of cereal chromosomes (ANANIEV *et al.* 1998; DONG *et al.* 1998; KISHII *et al.* 2001). However, no homology was found among the present 250-bp repeat and other tandem repeats. This poor conservation is contrary to the extensive presence of CCSI and pSau3A9 repeats in cereal species (ARAGÓN-ALCAIDE *et al.* 1996; JIANG *et al.* 1996). DNA sequences of CCS1 and pSau3AI are similar to those of the LTR and of the integrase gene, respectively, in Ty3/*gypsy*-like retroelements (MILLER *et al.* 1998; PRESTING *et al.* 1998). Although the DNA sequences of the 250-bp repeats showed no homology to *cereba* (PRESTING *et al.* 1998; HUDAKOVA *et al.* 2001) or any other retrotransposon-like sequences, their flanking or intervening sequences between the repeats showed high homology (>82%) to two separate sequences of *gag* gene and its upstream region (5′-UTR) in *cereba* (Figure 6A). This suggests that a number of *cereba*-like retroelements exist in the A and B genomes of wheat and that the 250-bp repeats are also related to *cereba*-like sequences. This suggestion was clearly supported by the 53% similarity obtained between the deduced amino acid sequences from the 250 bp with several CAAs and of the *gag* gene of *cereba* (Figure 6B), indicating that the DNA sequences

corresponding to the region [nucleotide position 1480–1925 in *cereba* (GenBank accession no. AY040832)] are highly divergent from barley to wheat.

Fukui *et al.* (2001) reported the presence of *cereba*-like retroelements in wheat centromeric regions, although they did not find the 250-bp tandem repeats or the region nonhomologous to *cereba*. So it is quite interesting to know whether *cereba*-like retroelements in wheat, named here *crew* (*c*entromeric *re*troelements of *w*heat), are still active or not. In most of our PCR products from wheat and *Ae. speltoides*, more than one stop codon appeared when their DNA sequences were translated into amino acids, but a few were uninterrupted. Like *cereba* in barley (Hudakova *et al.* 2001), therefore, complete and possibly autonomous Ty3/*gypsy*-like retroelements may be present in the genomes of wheat and/or it relatives.

As shown in Figure 6C, in the centromeric regions of wheat and *Ae. speltoides*, the 250-bp repeats with a few copies of CAA microsatellite are thought to be arrayed in tandem, but those with many copies are thought to be dispersed. This raises a question: Why is only the 250-bp unit amplified in tandem? Tandem repeats are common in centromeric regions of higher eukaryotes, and a number of amplification mechanisms have been suggested. The amplification of this tandem repeat was probably caused by the abundance of the *crew* sequences in the centromeric regions. We found highly conserved sequences (~40 and 48 bp, respectively; Figure 6C) at both the junctions of the 250- to 300-bp repeats and the sequences homologous to *cereba* in *crew*. Since they had no high homology to *cereba*, pBs301, or any other sequences in the DNA databases, their origin is uncertain. They might be important for amplifying the repeat in tandem, such as functioning as hot spots for recombination. Recombination might occur between a $(CCA)_2$ (GCA) (CCA) $(CAA)_2$(CCA) microsatellite array located at the end of the 48-bp junction sequence (Figure 6C) and the CAA microsatellite array in the 250- to 300-bp repeats.

Microsatellite variability is proposed to be generated by "DNA replication slippage," and to require a minimum number of repeats (reviewed by Schlötterer 2000). As described above, no DNA sequence homology between the region [nucleotide position 1480–1925 in *cereba* (accession no. AY040832)] and the 250-bp repeat was detected, but the enforced alignment between them revealed that the nucleotide sequence CAACCA$(CAA)_2$C in *cereba* (1630–1642 in accession no. AY040832) can be aligned with $(CAA)_4$C in pBS301-1 (96–108, Figure 5). This indicates that a $(CAA)_4$C sequence is a prototype of the microsatellite array. Both *cereba* in barley and *crew* in *Ae. speltoides* and wheat presumably originated from a single ancestral retrotransposon family, like *crwydryn* (Langdon *et al.* 2000). However, the 250-bp repeat of *crew* evolved to have many copies of CAA microsatellite within the sequence or popped out to form tandem arrays, while the corresponding sequences in *cereba* remained stable in the evolutionary process.

Tandem repeats are abundant components of centromere DNA in higher eukaryotes and have been thought to be related to centromere function, although their sequences are not homologous to each other. This study demonstrated the retroelement relationship of a novel centromeric 250-bp tandem repeat family and the possible origin and amplification mechanisms, suggesting that some tandem repeats localized at the centromeric regions of cereals and other plant species originated from parts of retrotransposons.

We thank Dr. J. S. (Pat) Heslop-Harrison for critical reading of this manuscript. This work was supported by the Core Research for Evolutionary Science and Technology Program and Japan Science Technology Corporation.

LITERATURE CITED

Ananiev, E. V., R. L. Phillips and H. W. Rines, 1998 Chromosomic-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc. Natl. Acad. Sci. USA **95:** 13073–13078.

Arabidopsis Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature **408:** 796–815.

Aragón-Alcaide, L., T. Miller, T. Schwarzacher, S. Reader and G. Moore, 1996 A cereal centromeric sequence. Chromosoma **105:** 261–268.

Bedbrook, J. R., J. Jones, M. O'Dell and R. B. Thompson, 1980 A molecular description of telemetric heterochromatin in *Secale* species. Cell **19:** 545–560.

Brandes, A., H. Thompson, C. Dean and J. S. Heslop-Harrison, 1997 Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. Chromosome Res. **5:** 238–246.

Brent, T. P., K. W. Michael and A. B. James, 2001 Characterization of a maize chromosome 4 centromeric sequence: evidence for evolutionary relationship with the B chromosome centromere. Genetics **159:** 291–302.

Chen, C. C., C. M. Chen, F. C. Hsu, C. J. Wang, J. T. Yang *et al.*, 2000 The pachytene chromosomes of maize as revealed by fluorescence *in situ* hybridization with repetitive DNA sequences. Theor. Appl. Genet. **101:** 30–36.

Cheng, Z., F. Dong, T. Langdon, S. Ouyang, C. R. Buell *et al.*, 2002 Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14:** 1691–1704.

Cheng, Z.-J., and M. Murata, 2002 Loss of chromosomes 2R and 5RS in octoploid triticale selected on agronomic characters. Genes Genet. Syst. **77:** 23–29.

Copenhaver, G. P., K. Nickel, K. Kuromori, M.-I. Benito, S. Kaul *et al.*, 1999 Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science **286:** 2468–2474.

Dong, F., J. T. Miller, S. A. Jackson, G.-L. Wang, P. C. Ronald *et al.*, 1998 Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc. Natl. Acad. Sci. USA **95:** 8135–8140.

Dvorak, J., and H.-B. Zhang, 1990 Variation in repeated nucleotide sequence sheds light on the phylogeny of the wheat B and G genomes. Proc. Natl. Acad. Sci. USA **87:** 9640–9644.

Friebe, B., and B. S. Gill, 1996 Chromosome banding and genome analysis in diploid and cultivated polyploid wheats, pp. 39–40 in *Methods of Genome Analysis in Plants*, edited by P. P. Jauhar. CRC Press, Boca Raton, FL.

Fukui, K. N., G. Suzuki, E. S. Lagudah, S. Rahman, R. Appels *et al.*, 2001 Physical arrangement of retrotransposon-related repeats in centromeric regions of wheat. Plant Cell Physiol. **42:** 189–196.

Hudakova, S., W. Michalek, G. G. Presting, R. ten Hoopen, K. Dos Santos *et al.*, 2001 Sequence organization of barley centromeres. Nucleic Acids Res. **29:** 5029–5035.

Jiang, J., S. Nasuda, F. Dong, C. W. Scherrer, S.-S. Woo *et al.*, 1996

A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc. Natl. Acad. Sci. USA **93:** 14210–14213.

Kamm, A., T. Schmidt and J. S. Heslop-Harrison, 1994 Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. Mol. Gen. Genet. **244:** 420–425.

Kishii, M., K. Nagaki and H. Tsujimoto, 2001 Tandem repetitive sequences located in the centromeric region of common wheat (*Triticum aestivum*) chromosomes. Chromosome Res. **9:** 417–428.

Langdon, T., C. Seago, M. Mende, M. Leggett, H. Thomas *et al.*, 2000 Retrotransposon evolution in diverse plant genomes. Genetics **156:** 313–325.

Miller, J. T., F. Dong, S. A. Jackson, J. Song and J. Jiang, 1998 Retrotransposon-related DNA sequence in the centromeres of grass chromosomes. Genetics **150:** 1615–1623.

Murata, M., 2002 Telomeres and centromeres in plants. Curr. Genomics **3:** 527–538.

Murata, M., N. Nakata and Y. Yasumuro, 1992 Origin and molecular structure of a midget chromosome in a common wheat carrying rye cytoplasm. Chromosoma **102:** 27–31.

Murata, M., Y. Ogura and F. Motoyoshi, 1994 Centromeric repetitive sequences in *Arabidopsis thaliana*. Jpn. J. Genet. **69:** 361–370.

Murata, M., J. S. Heslop-Harrison and F. Motoyoshi, 1997 Physical mapping of the 5S ribosomal RNA genes in Arabidopsis thaliana by multi-color fluorescence *in situ* hybridization with cosmid clones. Plant J. **12:** 31–37.

Nagaki, K., H. Tsujimoto and T. Sasakuma, 1998 A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions. Chromosome Res. **6:** 295–302.

Nonomura, K. I., and N. Kurata, 2001 The centromere composition of multiple repetitive sequences of rice chromosome 5. Chromosoma **110:** 284–291.

Pidoux, A. L., and R. C. Allshire, 2000 Centromere: getting a grip of chromosomes. Curr. Opin. Cell Biol. **12:** 308–319.

Presting, G. G., L. Malysheva, J. Fuchs and I. Schubert, 1998 A TY3/*GYPSY* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16:** 721–728.

Schlötterer, C., 2000 Evolutionary dynamics of microsatellite DNA. Chromosoma **109:** 365–371.

Schmidt, T., and J. S. Heslop-Harrison, 1996 The physical and genomic organization of microsatellites in sugar beet. Proc. Natl. Acad. Sci. USA **93:** 8761–8765.

Tsunewaki, K., 1996 Plasmonanalysis as the counterpart of genome analysis, pp. 271–299 in *Methods of Genome Analysis in Plants*, edited by P. P. Jauhar. CRC Press, Boca Raton, FL.

Communicating editor: J. A. Birchler