Characterization of a novel retrotransposon *TriRe-1* using nullisomic-tetrasomic lines of hexaploid wheat

Yuki Monden, Takeru Takai, Makoto Tahara
(*Course of Applied Plant Science*)

Retrotransposons constitute the large fraction (～80%) of the wheat genome where numerous and diverse retrotransposon families exist, where especially the long terminal repeat (LTR) retrotransposon family is known to be predominant. Thus, they have been considered to contribute to the genome expansion, sequence diversification and the genome structure alternation in the wheat genome. In addition, the insertion polymorphism of the LTR retrotransposon family among the cultivars has been known to be quite useful for the genetic analysis such as the linkage mapping and the phylogenetic studies. Here, we report the characteristics of a novel active LTR retrotransposon family *TriRe-1*, which belongs to the Ty1-*copia* group in the hexaploid wheat (*Triticum aestivum* L.) genome. This retroelement appears to encode all proteins required for the transposition and showed high insertion polymorphism among the hexaploid wheat cultivars, suggesting its potential of transpositional activity with at least recent transposition during wheat evolution. We studied the chromosomal localization of the *TriRe-1* insertion site based on the genome-wide comparative analysis using the nullisomic-tetrasomic lines of the cultivar Chinese Spring. The results showed that although the majority of the *TriRe-1* insertion sites exist across the homoeologous chromosomes of A, B or D genomes, a higher number of insertions in the B genome was detected compared to A or D genome, suggesting a specific amplification in the history of B genome progenitors. In conclusion, a novel LTR retrotransposon *TriRe-1* should be valuable for the development of molecular markers based on insertion polymorphism among the cultivars, and also the genome-specific *TriRe-1* insertion site can be utilized to study evolutional history of wheat genomes.

Key words : Retrotransposon, Wheat, Molecular markers, Nullisomic-tetrasomic lines

Introduction

Hexaploid bread wheat (2n=6x=42, AABBDD) is one of the most important cereal crops in the world and has a complex genome structure that derived from three different diploid species in the genera *Triticum* and *Aegilops*[5]. Two successive allopolyploidization events have led to the formation of this allohexaploid bread wheat. The first event generated the allotetraploid *T. turgidum* ssp. *durum* Hosn. (2n=4x=28, AABB) and occurred less than 0.5-0.6 MYA between the diploid species, *T. urartu* L. (2n=2x=14, AA) and an unidentified diploid species of the *Sitopsis* section which donated the B genome[5]. The second allopolyploidization occurred 7,000-12,000 years ago, and this hybridization was between the early domesticated tetraploid *T. dicoccon* Schrank and the diploid species *Ae. Tauchii* (2n=2x=14, DD), resulting in hexaploid wheat (2n=6x=42, AABBDD)[5]. The 21 pairs of homologous chromosomes of bread wheat fall into seven homoeologous groups, each group containing one pair of homologous chromosomes from the A, B and D genomes, respectively[21]. The homoeologous chromosomes, which are derived from the common ancestor, share a high degree of gene synteny and DNA sequence homology. Thus, the deleterious effects of nullisomic of a single pair of homologous chromosomes can be reduced or eliminated by making the tetrasomic condition of homoeologous chromosomes of one of the other two genomes. The nullisomic-tetrasomic wheat lines (cv Chinese Spring) were originally established by Sears et al[22]. This material has been used successfully to identify the chromosome for genome specific genes[2-3,7,12,15].

Retrotransposons are the mobile genetic elements that amplify their copy numbers within the genomes[13]. They are divided into long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons (LINEs and SINEs). The LTR retrotransposons have a similar structure to retrovirus, consisting of the LTR sequence at both their ends and the coding domains in the internal region. They encode the Gag, proteinase (PR), reverse transcriptase (RT)/RNaseH and integrase proteins (INT)[13]. They are subdivided into Ty1-*copia*

Received November 11, 2013
Graduate School of Environmental and Life Sciences, Okayama University, 1-1-1 Tsushimanaka, Kitaku, Okayama, Okayama 700-8502, Japan

or Ty3-*gypsy* members based on the order of their encoded proteins. Both are prevalent in the plant genome. Their replication process includes reverse transcription of RNA intermediates and the integration of newly replicated copies, leading to substantial accumulations of the copies in the genome[13].

In the wheat genome, the repetitive sequences cover the large fraction (~80%), wherein especially LTR retrotransposon families are known to be predominant[1,6,9]. Although there are a large number of LTR retrotransposon families in the plant genome, most retrotransposon families have already been inactivated[13]. Recently, we successfully cloned the DNA segment of a novel active retrotransposon *TriRe-1* in the wheat genome. This family showed high insertion polymorphism among the closely related cultivars that have been developed in Japan, which suggested that this family should have some transpositional activity or at least recent transposition[25].

With the development of high-throughput sequencing technology such as Roche 454 and Illumina HiSeq 2000, a large amount of sequencing data can be obtained in a short time. Retrotransposons have numerous copies (several hundreds or thousands) in the genome, which require a large amount of sequencing data for the genome-wide screening of their insertion sites. Thus, high-throughput sequencing platform seemed to be a suitable tool for screening these multiple retrotransposon insertion sites[31]. In fact, there are several reports that have sequenced a large number of retrotransposon insertion sites with a high-throughput sequencing platform[4,11,28].

In this research, we revealed the molecular characteristics of a novel active retrotransposon family *TriRe-1* and identified their extensive insertion sites using diterosomic and nullisomic-tetrasomic lines with HiSeq2000 sequencing platform. Out of these insertion sites, we focused on the insertion sites that were detected in all lines except for a single diterosomic or nullisomic-tetrasomic line to reveal their chromosomal localization in the wheat genome. Our results showed that *TriRe-1* has more genome specific insertions in the B genome than in the A and D genomes, which indicated that this family might have amplified their copy numbers specifically in the B genome.

Materials and Methods

DNA samples

Young leaves were collected from the wheat cultivar Chinese Spring, Hokushin, 2 diterosomic lines and 35 nullisomic-tetrasomic lines among the possible 42 tetranullisomic combinations. Genomic DNAs were extracted with the DNeasy Plant mini kit (QIAGEN) according to the manufacturer's protocol.

Cloning of the entire region of TriRe-1

In the previous research, we detected one *TriRe-1* insertion site that was specific to Hokushin cultivar[25]. We determined the sequence of this insertion site from another wheat cultivar and designed the PCR primers to clone the entire region of the *TriRe-1* inserted copy (Fig. 1A). The PCR amplification was performed using SB-RT90_650 primer and Hoku3 primer sets. We excised this PCR product from 1.5% of agarose gel and purified with QIAquick Gel Extraction kit (QIAGEN). The purified product was cloned using TOPO® XL Cloning Kit (Invitrogen). We tried to sequence the entire insert of the TOPO XL clone by genome walking but experienced difficulties. Therefore, we randomly fragmented the excised PCR product with a nebulizer (Invitrogen) and the fragments were sub-cloned by the TOPO TA Cloning® Kit (Invitrogen). The overlapping sequences of the randomly broken fragments were assembled to form entire inserts. To confirm the sequence accuracy of the entire insert, we sequenced enough fragments that any base position of the assembled insert was sequenced by at least three different fragments. The sequence information of primes used here was listed in Supplemental Table 1.

Characterization and annotation of TriRe-1

The characteristics of *TriRe-1* sequence was studied for the cis-element by CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html), the ORF region by ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf) and the conserved motif by the Protein Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and HMMER (http://hmmer.janelia.org/).

Preparation of Illumina NGS sequencing libraries

Genomic DNAs were fragmented with nebulizer and purified with QIAquick PCR Purification Kit (QIAGEN). The purified products were end-repaired, modified to add 3'A overhangs, and ligated to the forked adapters. The ligated products underwent primary amplification using RNaseH *TriRe-1* primer and LAM-AP2 HQ primer sets and secondary amplification using the *TriRe-1*_LTR_Down primer and LAM-AP3 HQ primer combinations. The 5-8 bp barcode sequences were added to the *TriRe-1*_LTR_Down

primer for wheat line identification. The PCR products were size-selected (300-500bp) by gel electrophoresis and purified with QIAquick Gel Extraction Kit (QIAGEN). Those PCR products were pooled into one sequencing sample. This sequencing sample was qualified by Bioanalyzer (Agilent Technologies, Inc., Santa Clara, California, USA). Paired-end sequencing reads were generated on an Illumina HiSeq2000 platform. The sequence information of adaptor and primer were listed in Supplemental Table 1.

Data analysis

Sequencing data were handled on the analysis pipeline execution system of the Cell Innovation program at the National Institute of Genetics (NIG), which is named Maser (http://cell-innovation.nig.ac.jp/index_en.html). Read pairs were filtered and assigned by *TriRe-1* sequence with 5-8bp barcode sequences of the cultivar (Table 1 and Supplemental Table 2). When one read of a sequence pair was filtered for invalid *TriRe-1* or barcode sequence, the entire pair was discarded. After the *TriRe-1* sequence with 5-8bp barcode sequences was deleted from the reads, we trimmed these reads to the specific length with which over 99% of sequences were included, and shorter reads than this specific length were filtered out (Table 1). The resulting sequences were trimmed to 50 bp from the *TriRe-1* junction, and filtered with quality-value of all base calls being equal or above 30 (Table 1). Subsequently, the identical sequences over 10 reads were collapsed into a single sequence in the fasta format with read counts information, and less than 10 identical sequences were discarded. Those sequences were used for the clustering analysis with blat based on the sequence similarity (using the following parameters : -tileSize 8, -minMatch 1, -minScore 10, -repMatch -1, -oneOff 2)[10]. Each cluster of sequences or non-clustered sequences with more than 10 reads should represent an individual insertion site where a copy of *TriRe-1* inserted in at least one or more cultivars. However, assigning the insertion site into cultivars is likely to involve some errors ; sequence errors of the index sequence may allocate the read to an incorrect cultivar and the sequence with small read number may be mistakenly included in the large cluster by the blat analysis. These erroneous assignments should result in a very small number of the reads. For this reason, we set the critical value to judge the presence of the insertion for *TriRe-1* ; if the reads of a cultivar at the particular insertion site is less than 0.01% of the entire reads for that cultivar, we presumed the retrotransposon was absent at that site. The genotyping was conducted based on the presence (1) or absence (0) of information at each site over all insertion sites. The number of sequence reads that were further assigned to each cultivar is shown in Supplemental Table 2.

PCR confirmation of identified insertion sites

We designed the PCR primer based on the genomic sequence of the wheat line-specific insertion site (Fig. 3A, Supplemental Table 1). The PCR amplification was performed using the line-specific and *TriRe-1* primer pair (Fig. 3A, 3B). We also designed the PCR primer based on the junction sequence of *TriRe-1* insert to reveal whether there were any genomic rearrangements or mutations existing around the *TriRe-1* insert (Fig. 3A, 3B). The information regarding primer sequences is listed in Supplemental Table 1.

Results

Molecular characterization of a novel retrotransposon TriRe-1

In the previous research, we identified one *TriRe-1* insertion site which was specific to the Hokushin cultivar[25]. From the long PCR amplification of this Hokushin specific insertion site, the 5,432 bp PCR product was obtained (Fig. 1A). We extracted and fragmented this PCR product to reveal the *TriRe-1* entire sequence. A total of 32 DNA fragment sequences were obtained and assembled to cover the *TriRe-1* entire region (Fig. 1B). As a result, the length of *TriRe-1* entire region was estimated to be 5,267bp including 116bp of 5'LTR and 117bp of 3'LTR sequences (Fig. 1A and 1B). Its insertion site is flanked by a 5bp target site duplication (TSD) (Fig. 1A). In the internal region, the primer binding site (PBS) and polypurine tract (PPT) sequences were detected adjacent to 5'LTR and 3'LTR regions, respectively (Fig. 1C). The coding region was consisted of 1,636 amino-acid sequences which contained all proteins needed for transposition including gag, protease, integrase, reverse transcriptase and RNaseH (Fig. 1C). This element was considered to be a Ty1-copia family member because the motif sequences of this element have extremely high similarity with that of known Ty1-*copia* elements (CIRE1[20], *Tto1*[26], *BARE-1*[16], *Tnt1*[29], *Tlc1*[27], *copia*[30], *FaRE1*[18], *Rtsp-1*[23,24]) (Fig. 1C). The 8bp of PBS sequence (TGGTATCA) was consistent with that of other elements (CIRE1, *Tto1*, *Tnt1*, *Tlc1*, *FaRE1* and *Rtsp-1*), assuming the reverse transcription of this element is also

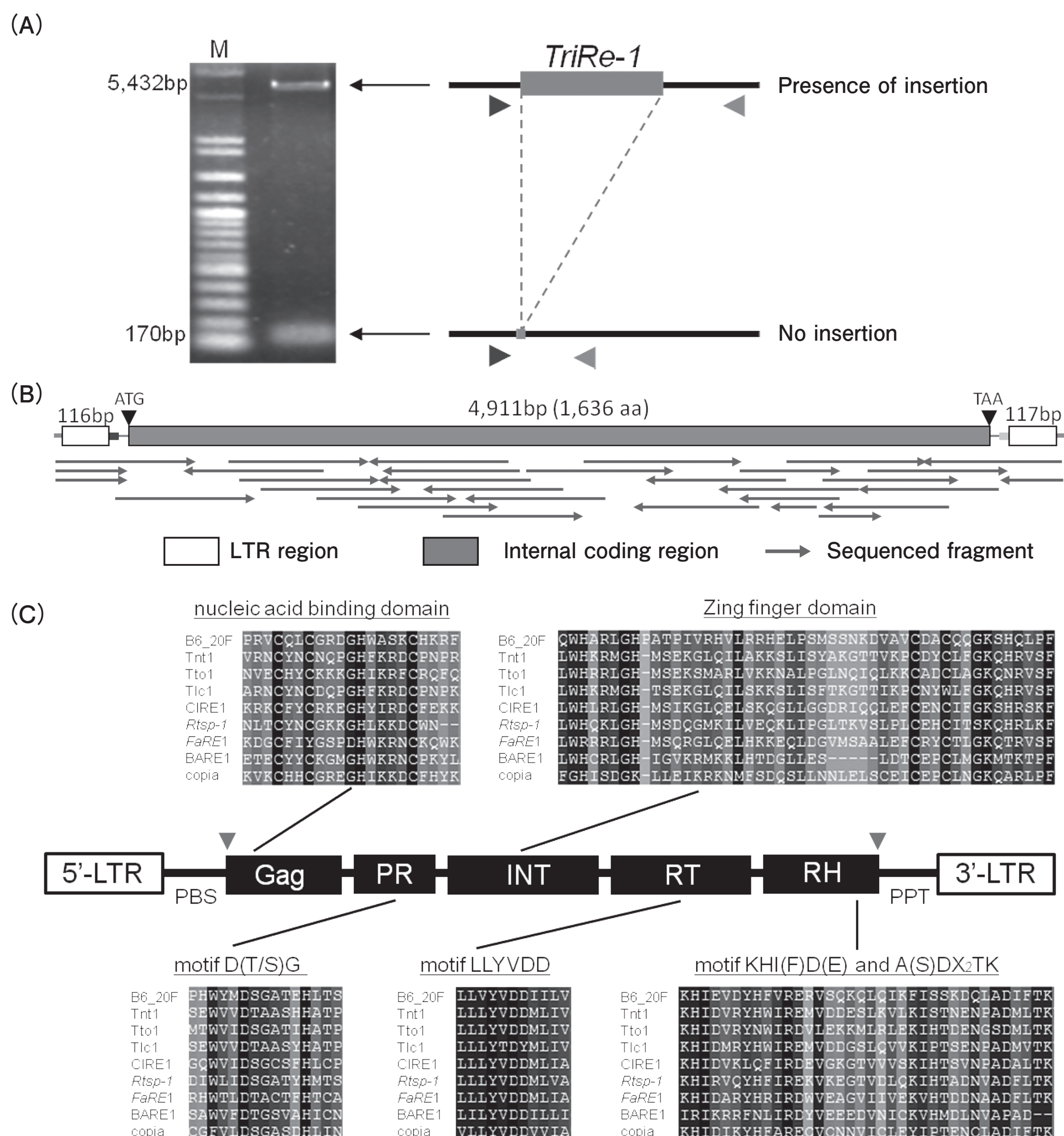

Fig. 1 Structure of the *TriRe-1* element. (A) Long PCR using PCR primers (represented by arrows) designed at flanking regions of *TriRe-1*. The electrophoresis was conducted on 1% agarose gel. The length of the long band is 5,432bp and that of the short band is 170bp. (B) Determination of the entire sequence of *TriRe-1*. A total of 32 DNA fragments (represented by arrow) were sequenced and connected to cover the entire region. (C) The whole structure of a *TriRe-1* element and the alignment of amino acid sequences with other well-characterized Ty1-*copia* elements in protein coding regions.

initialized with priming iMET tRNA with this PBS site.

Identification of TriRe-1 insertion sites with HiSeq2000 sequencing

We constructed a sequencing library to identify *TriRe-1* insertion site extensively in the wheat cultivar Chinese Spring, Hokushin, two diterosomic lines and 35 nullisomic-tetrasomic lines. PCR amplification was performed using *TriRe-1* specific primer which contains the 5-8bp barcode sequences for discriminating the cultivar. These PCR products were pooled into a single sequencing sample and sequenced with HiSeq2000 sequencing platform. After sequencing, a total of 155,488,208 of 101bp paired reads were obtained ; 86.2% of which (134,024,377 paired reads) contained the exact *TriRe-1* and barcode sequences, which indicated that we could specifically amplify the DNA fragments at the *TriRe-1* insert site (Table 1). After trimming and filtering based on the quality value of all base calls being ≥ 30 (Materials and Methods), 47.3% of them (73,554,167 paired reads) remained, which

produced a total of 78,509 different sequences, each of which has at least 10 reads (Table 1). We conducted the clustering analysis using these 78,509 sequences with the blat application (using the following parameters : -tileSize 8, -minMatch 1, -minScore 10, -repMatch-1, -oneOff 2) based on the sequence similarity. The clustering analysis resulted in a total of 5,797 *TriRe-1* insertion sites (2,990 clusters and 2,807 non-clustered single sequences) (Table 1). We analyzed these insertion data to determine the presence (1) and absence (0) of *TriRe-1* insertion at each insertion site for all lines based on the criteria (Materials and Methods). We also created a matrix sheet to compare the *TriRe-1* insertion sites among the lines. The result showed that the average number of *TriRe-1* insertion site per cultivar was estimated to be 476 (data not shown).

Table 1 Summary of sequence reads in the data processing

Analysis	No. of reads	No. of collapsed reads (>=10)	Ratio(%)	No. of clusters
Raw data	155,488,208	—	—	—
TriRe-1 sequence valid	134,024,377	327,735	86.2	—
Trimming to 50bp	133,926,509	135,450	86.1	—
QV Filtering (>=30)	73,554,167	78,509	47.3	—
Clustering with blat	—	—	—	5,797

The chromosomal localization analysis of TriRe-1 insertion site using the nullisomic-tetrasomic lines

The nullisomic-tetrasomic lines of bread wheat cultivar Chinese Spring have been successfully used to localize genome specific genes on particular chromosomes of hexaploid wheat[2-3,7,12,15]. This type of line is lacking a pair of chromosomes that is replaced by an extra pair of homoeologous chromosomes. In other words, these lines are nullisomic (N) for one pair of the homoeolog chromosomes, and tetrasomic (T) for one of the two pairs of other homoeolog chromosomes. Thus, the chromosomal localization of *TriRe-1* insertion site could be determined by comparing the insertion sites among these nullisomic-tetrasomic lines. For example, a *TriRe-1* insertion site was observed in all lines except in the line NT3B3D, which strongly indicates this insertion site should be localized on chromosome 3B (Fig. 2A). Among the 42 tetra-nullisomic combinations possible, 35 combinations were obtained and have been tested. The nullisomic 2A and 4B lines were not obtained. Overall 14, 46 and one *TriRe-1* insertion sites were considered to be localized specifically on the A, B and D genomes, respectively (Fig. 2A). These insertion sites of A and B genomes were distributed throughout all seven chromosomes, and the number of insertions per homoeologous chromosome ranged from two on chromosome 3 to 17 on chromosome 6 (Fig. 2B). Out of these 61 insertion sites, we randomly selected 14 insertion sites to confirm them by PCR-

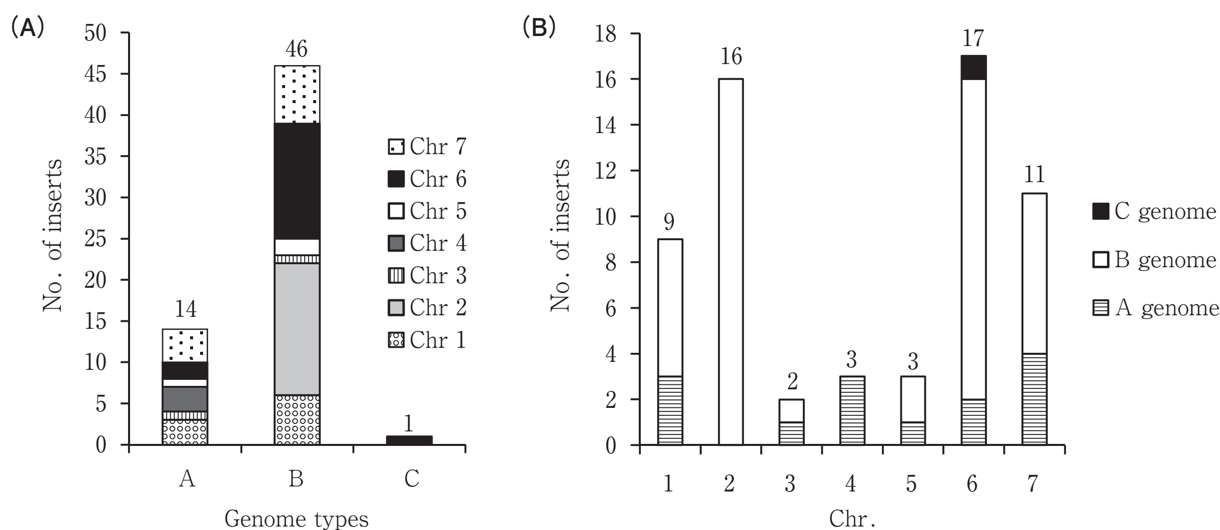

Fig. 2 The number of *TriRe-1* insertion sites identified in the data analysis. (A) Showing the number of *TriRe-1* insertion sites in each genome. The B genome has a higher number of *TriRe-1* insertion sites. (B) Showing the number of *TriRe-1* insertion sites in each chromosome.

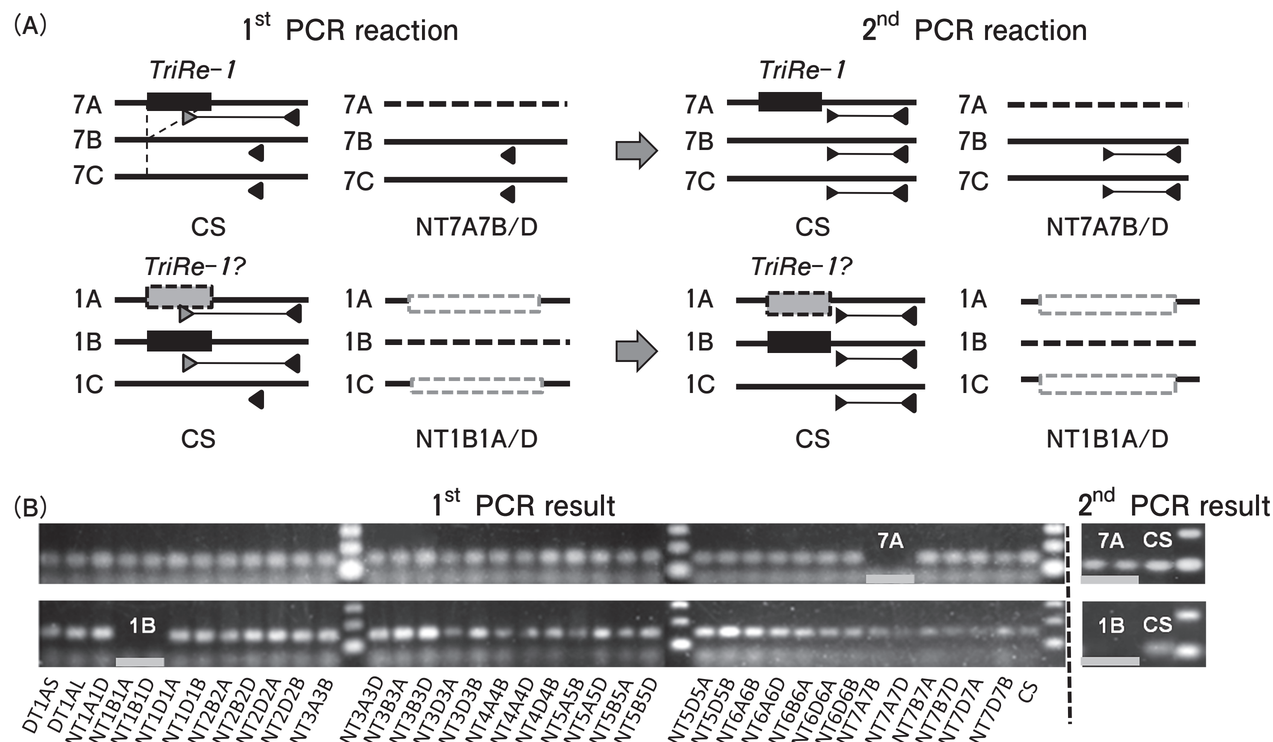

Fig. 3 PCR verification of the *TriRe-1* insertion site identified in nullisomic-tetrasomic lines. (A) The scheme of the PCR analysis. At the 1st PCR reaction, we used *TriRe-1* specific (gray arrow) and insertion site specific (black arrow) primer sets. Next, we used insertion site specific primer (black arrows) sets to investigate whether there are any other genomic rearrangements at the *TriRe-1* insert site or not. (B) The image of agarose gel electrophoresis. The upper gel showed all lines except for NT7A7B/D amplified the DNA fragments at the 1st PCR reaction. Also, Chinese Spring and NT7A7B/D showed the same DNA fragment at the 2nd PCR reaction. These results indicated that a *TriRe-1* insert should be located at 7A chromosome. In contrast, the lower gel showed that although Chinese Spring showed the DNA fragment, NT1B1A/D didn't amplify the DNA fragment at the 2nd PCR reaction. This result indicated that there might be some genomic rearrangement (gray dotted box) at the *TriRe-1* insert region in this line. In this case, we couldn't determine whether a *TriRe-1* insert was located at 1B chromosome or not.

based analysis; the first PCR reaction was performed to validate the existence of a *TriRe-1* insertion by designing the PCR primers at the *TriRe-1* specific and the insertion site specific sequences (Fig. 3A, Supplemental Fig. 1A, Table 2), and the second PCR reaction was performed to investigate whether there are any other genomic rearrangement/mutation except for a *TriRe-1* insertion at the insert site in the target line (Fig. 3A, Supplemental Fig. 1B). These approaches led to the conclusion that 85.7% of these insertions (12/14) were confirmed and located to be specifically in each chromosome (Fig. 3B, Supplemental Fig. 1B, Table 2). We determined the chromosomal localization of seven, four and one *TriRe-1* insertion sites that were specific in the A, B and D genomes, respectively. By contrast, at the remaining two (1B and 6B) insertion sites, other genomic rearrangement such as insertion/deletion or nucleotide substitution might exist, because we couldn't amplify the DNA fragment using insert site specific primer at the 2nd PCR reaction (Fig. 3B, Supplemental Fig. 1B, Table 2). In this case, we couldn't determine the localization of a *TriRe-1*. Since the total number of the *TriRe-1* insertion site was estimated to be 476 on average in the wheat genome, most insertion sites should exist across the homoeologous chromosomes. Thus, the majority of *TriRe-1* insertion sites should have existed in the common ancestors. Also, after the divergence from the common ancestors, they might have amplified their copy number specifically in a diploid progenitor of the B genome.

Discussion

In this research, we revealed the characteristics of a novel active LTR retrotransposon family *TriRe-1* in the wheat genome, and extensively identified their insertion sites using the nullisomic-tetrasomic lines with

Table 2 Summary of the PCR analysis for investigating the chromosomal location of insertion site

Chr.	Line/cultivar	Putative chromosomal location													
		1AS	1AL	1B*	2B	3A	3B	4A	5A	5B	6A	6B*	6D	7A	7B
		Cluster name													
		pattern 345	pattern 400	pattern 578	pattern 408	pattern 101	pattern 52	pattern 369	pattern 98	pattern 449	pattern 149	pattern 311	pattern 129	pattern 267	pattern 562
1	DT1AS	−	+	+	+	+	+	+	+	+	+	+	+	+	+
	DT1AL	+	−	+	+	+	+	+	+	+	+	+	+	+	+
	NT1A1D	−	−	+	+	+	+	+	+	+	+	+	+	+	+
	NT1B1A	+	+	−	+	+	+	+	+	+	+	+	+	+	+
	NT1B1D	+	+	−	+	+	+	+	+	+	+	+	+	+	+
	NT1D1A	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	NT1D1B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
2	NT2B2A	+	+	+	−	+	+	+	+	+	+	+	+	+	+
	NT2B2D	+	+	+	−	+	+	+	+	+	+	+	+	+	+
	NT2D2A	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	NT2D2B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
3	NT3A3B	+	+	+	+	−	+	+	+	+	+	+	+	+	+
	NT3A3D	+	+	+	+	−	+	+	+	+	+	+	+	+	+
	NT3B3A	+	+	+	+	+	−	+	+	+	+	+	+	+	+
	NT3B3D	+	+	+	+	+	−	+	+	+	+	+	+	+	+
	NT3D3A	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	NT3D3B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4	NT4A4B	+	+	+	+	+	+	−	+	+	+	+	+	+	+
	NT4A4D	+	+	+	+	+	+	−	+	+	+	+	+	+	+
	NT4D4B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
5	NT5A5B	+	+	+	+	+	+	+	−	+	+	+	+	+	+
	NT5A5D	+	+	+	+	+	+	+	−	+	+	+	+	+	+
	NT5B5A	+	+	+	+	+	+	+	+	−	+	+	+	+	+
	NT5B5D	+	+	+	+	+	+	+	+	−	+	+	+	+	+
	NT5D5A	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	NT5D5B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
6	NT6A6B	+	+	+	+	+	+	+	+	+	−	+	+	+	+
	NT6A6D	+	+	+	+	+	+	+	+	+	−	+	+	+	+
	NT6B6A	+	+	+	+	+	+	+	+	+	+	−	+	+	+
	NT6D6A	+	+	+	+	+	+	+	+	+	+	+	−	+	+
	NT6D6B	+	+	+	+	+	+	+	+	+	+	+	−	+	+
7	NT7A7B	+	+	+	+	+	+	+	+	+	+	+	+	−	+
	NT7A7D	+	+	+	+	+	+	+	+	+	+	+	+	−	+
	NT7B7A	+	+	+	+	+	+	+	+	+	+	+	+	+	−
	NT7B7D	+	+	+	+	+	+	+	+	+	+	+	+	+	−
	NT7D7A	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	NT7D7B	+	+	+	+	+	+	+	+	+	+	+	+	+	+
−	Chinese Spring	+	+	+	+	+	+	+	+	+	+	+	+	+	+

*These insertion sites should have other genomic rearrangements, apart from *TriRe-1* insertion.

HiSeq 2000 sequencing platform. We conducted a comparative analysis of these insertion sites among nullisomic-tetrasomic lines to reveal their chromosomal location. Our results showed that although the majority of *TriRe-1* insertion sites were found across the homoeologous chromosomes, the B genome has a higher number of *TriRe-1* genome specific insertion site than A and D genomes.

To date, there are several reports that have investigated the copy number of a retrotransposon family in A, B and D genomes using nullisomic-tetrasomic lines[12], the wild putative progenitors (*Aegilops speltoides* Taush (diploid, genome B), *Ae. Tauschii* Cross (diploid, genome D), *Ae. Sharonensis* Eig (diploid, genome S) and *T. urartu* L. (diploid, genome A)), and the cultivated species (*T. monococcum* L. (diploid, genome A), *T. dicoccon* Schrank (tetraploid, genome AB), and *T. spelta* L. (hexaploid, genome ABD))[17]. For example, Pagnotta et al.[17] suggested that a Ty1-*copia* member namely, WIS2-1A retroelement has more insertion sites in the B genome than the A and D genomes. Konovalov et al.[12] reported that a Ty3-*gypsy* member namely, *Jeli* retroelement is predominantly abundant in the A genome compared to B and D genomes, in contrast, a Ty1-*copia* member namely, *BARE-1* has similar copy numbers in the A and B genomes with a reduced number in the D genome. In our case, a *TriRe-1* retroelement has a higher copy number in the B genome compared to A and D genomes. Considering these results, different LTR retrotransposon families probably proliferated and distributed at different rates and in the different genomes due to an independent evolution of each LTR retrotransposon family. These proliferations of specific LTR retrotransposon families in the specific genome might contribute to the genome expansion, sequence divergence and the genome size variation in the wheat genome.

Retrotranposon based molecular markers, namely, sequence-specific amplification polymorphism (SSAP), retrotransposon microsatellite amplification polymorphism (REMAP), and inter-retrotransposon amplification polymorphism (IRAP) have been developed in many crop species for the linkage analysis, cultivar or species identification and the phylogenetic studies[8,14,19]. The *TriRe-1* retroelement has also the potential to be a quite useful source as a novel molecular marker because of its recent activity and the associated insertion site polymorphism among the wheat cultivars. Moreover, the genome specific *TriRe-1* insertion site identified in this study should be useful to study evolutional history of wheat genomes.

Acknowledgements

We thank Dr. Nobuyuki Fujii and Dr. Kazuho Ikeo in the National Institute of Genetics for the development of data analysis pipelines. This work was supported by the Program to Disseminate Tenure Tracking System, from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan (to Y.M) and cooperative research program with Nippon Flour Mills (NIPPN).

References

1) Bendich, A. J. and B. J. McCarthy : DNA comparisons among barley, oats, rye, and wheat. Genetics., **65**, 545-565 (1970)
2) Bittner, T., S. Campagne, G. Neuhaus, S. A. Rensing and C. Fischer-Iglesias : Identification and characterization of two wheat Glycogen Synthase Kinase 3/SHAGGY-like kinases. BMC Plant Biol., **13**, 64 (2013)
3) Brewer, G. J., C. F. Sing and E. R. Sears : Studies of isozyme patterns in nullisomic-tetrasomic combinations of hexaploid wheat. Proc. Natl. Acad. Sci. USA., **64**, 1224-1229 (1969)
4) Ewing, A. D. and H. H. Kazazian : Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. Genome Res., **21**, 985-990 (2011)
5) Feldman, M. and A. A. Levy : Genome evolution due to allopolyploidization in wheat. Genetics., **192**, 763-774 (2012)
6) Flavell, R. B., J. Rimpau and D. B. Smith : Repeated sequence DNA relationship in four cereals genomes. Chromosoma., **63**, 205-222 (1977)
7) Joppa, L. R., K. Khan and N. D. Williams : Chromosomal location of genes for gliadin polypeptides in durum wheat *Triticum turgidum* L. Theorl. Appl. Genet., **64**, 289-293 (1983)
8) Kalendar, R., A. J. Flavell, T. H. N. Ellis, T. Sjakste, C. Moisy and A. H. Schulman : Analysis of plant diversity with retrotransposon-based molecular markers. Heredity., **106**, 520-530 (2011)
9) Keller, B. and C. Feuillet : Colinearity and gene density in grass genomes. Trend Plant Sci., **5**, 246-251 (2000)
10) Kent, W. J. : BLAT—the BLAST-like alignment tool. Genome Res., **12**, 656-664 (2002)
11) Kofler, R., A. J. Betancourt and C. Schlötterer : Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. PLoS Genet., **8**, e1002487 (2012)
12) Konovalov, F. A. and N. P. Goncharov : Molecular markers based on LTR retrotransposons *BARE-1* and *Jeli* uncover different strata of evolutionary relationships in diploid wheats. Mol. Genet. Genomics., **283**, 551-563 (2010)
13) Kumar, A. and J. L. Bennetzen : Plant retrotransposons. Annu. Rev. Genet., **33**, 479-532 (1999)
14) Kumar, A. and H. Hirochika : Applications of retrotransposons as genetic tools in plant biology. Trends Plant Sci., **6**, 127-134 (2001)

15) Li, A., W. Yang, S. Li, D. Liu, X. Guo, J. Sun and A. Zhang : Molecular characterization of three *GIBBERELLIN-INSENSITIVE DWARF1* homologous genes in hexaploid wheat. J Plant Physiol., **170,** 432-443 (2013)

16) Manninen, I. and A. H. Schulman : *BARE-1, a copia-like* retroelement in barley (*Hordeum vulgare* L.). Plant Mol. Biol., **22,** 829-846 (1993)

17) Pagnotta, M. A., L. Mondini and E. Porceddu : Quantification and organization of WIS2-1A and BARE-1 retrotransposons in different genomes of *Triticum* and *Aegilops* species. Mol. Genet. Genomics., **282,** 245-255 (2009)

18) Ping, He., M. Yue, Z. Guiling, D. Hongyan, L. He, C. Linlin and Z. Zhang : *FaRE1* : a transcriptionally active Ty1-*copia* retrotransposon in strawberry. J. Plant Res., **123,** 707-714 (2011)

19) Poczai, P., I. Varga, M. Laos, A. Cseh, N. Bell, J. P. Valkonen and J. Hyvönen : Advances in plant gene-targeted and functional markers : a review. Plant Methods., **9,** 6 (2013)

20) Rico-Cabanas, L. and J. A. Martı'nez-Izquierdo : CIRE1, a novel transcriptionally active Ty1-*copia* retrotransposon from *Citrus sinensis*. Mol. Genet. Genomics., **277,** 365-377 (2007)

21) Sears, E. R. : The aneuploids of common wheat. Missouri Agric. Exp. Stn. Res. Bull. **572,** 1-58 (1954)

22) Sears, E. R. : Nullisomic-tetrasomic combinations in hexaploid wheat. In Chromosome Manipulation and Plant Genetics., 29-45 (1966)

23) Tahara, M., T. Aoki, S. Suzuka, H. Yamashita, M. Tanaka, S. Matsunaga and S. Kokumai : Isolation of an active element from a high-copy-number family of retrotransposons in the sweetpotato genome. Mol. Gen. Genomics., **272,** 116-127 (2004)

24) Tahara, M. and H. Yamashita : A Novel Initiation Complex for Reverse Transcription of an Active LTR Retrotransposon in Sweetpotato. Scientific Reports of the Faculty of Agriculture Okayama University., **96,** 7-11 (2007)

25) Takai, T. and M. Tahara : Discovery of the retrotransposon showing genome insertion polymorphisms among wheat cultivars. DNA polymorphism., **20,** 80-90 (2012)

26) Takeda, S., K. Sugimoto, H. Otsuki and H. Hirochika : Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyl jasmonate. Plant Mol. Biol., **36,** 365-376 (1998)

27) Tapia, G., I. Verdugo, M. Yanez, I. Ahumada, C. Theoduloz, C. Cordero, F. Poblete, E. Gonzalez and S. Ruiz-Lara : Involvement of ethylene in stress-induced expression of the *TLC1-1* retrotransposon from *Lycopersicum chilense* Dun. Plant Physiol., **138,** 2075-2086 (2005)

28) Urbański, D. F., A. Małolepszy, J. Stougaard and S. U. Andersen : Genome-wide *LORE1* retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. Plant J., **69,** 731-741 (2012)

29) Vernhettes, S., M. A. Grandbastien and J. M. Casacuberta : In vivo characterization of transcriptional regulatory sequences involved in the defence-associated expression of the tobacco retrotransposon *Tnt1*. Plant Mol. Biol., **35,** 673-679 (1997)

30) Yoshioka, K., H. Kanda, N. Takamatsu, S. Togashi, S. Kondo, T. Miyake, Y. Sasaki and T. Shiba : Efficient amplification of Drosophila simulans *copia* directed by high-level reverse transcriptase activity associated with *copia* virus-like particles. *Gene*., **120,** 191-196 (2012)

31) Xing, J., D. J. Witherspoon and L. B. Jorde : Mobile element biology : new possibilities with high-throughput sequencing. Trends Genet., **29,** 280-289 (2013)

コムギ染色体欠損系統を用いた新規活性型レトロトランスポゾン *TriRe-1* の分子遺伝学的解析

門田　有希・高井　　健・田原　　誠
（応用植物科学コース）

レトロトランスポゾンは植物ゲノムの主要な構成要素であり，コムギゲノムにおいてはその80％を占める．特にLTR 型レトロトランスポゾンの割合が高く，ゲノムの拡大，配列の多様性およびゲノム構造変異等に大きく寄与してきたと考えられている．これら配列は自身のコピー配列を複製し増幅するため，ゲノム中には数百，数千に及ぶコピー配列をもつ．また，ゲノム進化の過程において多数のファミリーを形成してきた．これら多数のファミリーのうち，現在でも転移活性を示す活性型ファミリーは，品種間において高い挿入多型を示すことが知られている．このような挿入多型は，連鎖解析および系統解析等各種遺伝解析に利用可能である．本研究では，コムギにおける新規活性型レトロトランスポゾンファミリー *TriRe-1* の特徴を詳細に解析した．*TriRe-1* は転移に必要なタンパク質をコードする内部配列をもち，また日本で育成されたコムギ近縁品種間においても高い挿入多型を示したため，現在でも転移活性を有している，もしくはごく最近まで転移していた可能性が高いと考えられた．一方で，コムギ染色体欠損系統（ナリソミックテトラソミック系統）を用い，*TriRe-1* の挿入箇所を比較解析した．その結果，大部分の挿入箇所は複数の同祖染色体に存在すると考えられたが，Ｂゲノムにおいて最も多くの特異的な挿入箇所が同定された．よって，Ｂゲノム祖先種において活発に増幅してきた可能性が示唆された．今回の結果により，新規活性型レトロトランスポゾン *TriRe-1* の品種間挿入多型を利用した DNA マーカー，また，各ゲノム（A，B，Dゲノム）特異的な挿入箇所を利用したゲノム識別性に優れた DNA マーカーの開発の可能性が期待される．

Received November 11, 2013
岡山大学大学院環境生命科学研究科