Genes Genet. Syst. (2008) **83**, p. 13–22

The pattern of amplification and differentiation of *Ty1-copia* and *Ty3-gypsy* retrotransposons in Brassicaceae species

Ryo Fujimoto, Shohei Takuno, Taku Sasaki and Takeshi Nishio*

Graduate School of Agricultural Science, Tohoku University, Aoba-ku, Sendai 981-8555, Japan

(Received 3 September 2007, accepted 21 November 2007)

One of the causes of genome size expansion is considered to be amplification of retrotransposons. We determined nucleotide sequences of 24 PCR products for each of six retrotransposons in *Brassica rapa* and *Brassica oleracea*. Phylogenetic trees of these sequences showed species-specific clades. We also sequenced *STF7a* homologs and *Tto1* homologs, 24 PCR products each, in nine diploids and three allopolyploids, and constructed phylogenetic trees. In these phylogenetic trees, species-specific clades of diploid species were also formed, but retrotransposons of allopolyploids were clustered into the clades of their original genomes, indicating that these two retrotransposons amplified after speciation of the nine diploids. Genetic variation in these retrotransposons may have arisen before emergence of allopolyploid species. There was a positive correlation between the genome size and the average number of substitutions of *STF7a* and *Tto1* homologs in at least seven diploids. The implications of these results in the genome evolution of Brassicaceae are herein discussed.

Key words: Brassicaceae, *Ty1-copia*, *Ty3-gypsy*, phylogenetic analysis, genome size, sequence diversity

INTRODUCTION

Retrotransposons are ubiquitous in the plant kingdom (Kumar and Bennetzen, 1999), but those still having transposing activity, such as *Tos17*, *Tto1*, *Tnt1*, and *LORE1* (Grandbastien et al., 1989; Hirochika, 1993; Hirochika et al., 1996; Madsen et al., 2005), are few in plants. Retrotransposons can be divided into two groups, *i.e*., LTR (long terminal repeat) retrotransposons and non-LTR retrotransposons, and the LTR retrotransposons are classified into *Ty1-copia* and *Ty3-gypsy* (Xiong and Eickbush, 1990). The copy number of retrotransposons tends to increase because of their copy-and-paste mode of transposition. Retrotransposons have been estimated to constitute approximately 80% of the genome in maize, which has a large genome size (SanMiguel et al., 1998), but less than 10% in *Arabidopsis thaliana,* which has a small genome size (Arabidopsis Genome Initiative, 2000). Transposition and amplification of retrotransposons are considered to play a major role in the evolution of plant genomes.

Brassicaceae consists of about 340 genera and more than 3,350 species including important crops and a model plant, *A. thaliana* (Al-Shehbaz, 1984). Important crop species, such as *Brassica oleracea*, *Brassica rapa*, *Brassica napus*, and *Raphanus sativus,* belong to the tribe Brassiceae. *Brassica juncea, Brassica carinata,* and *B. napus* are allopolyploids having AABB, BBCC, and AACC genomes, respectively, the A, B, and C genomes of which are derived from *B. rapa*, *Brassica nigra,* and *B. oleracea,* respectively (U, 1935). Brassicaceae is an interesting family for the study of genome evolution because of its large variations of chromosome number and genome size (Arumuganathan and Earle, 1991; Johnston et al., 2005).

Many species in Brassicaceae have self-incompatibility, which is controlled by the *S* locus (Bateman, 1955). About 50 and 30 *S* haplotypes have been identified in *B. oleracea* and *B. rapa*, respectively (Nou et al., 1993; Ockendon, 2000). Interspecific pairs of *S* haplotypes having the same recognition specificity have been identified between *B. rapa* and *B. oleracea* (Kimura et al., 2002; Sato et al., 2003). Comparison of the genome structure of the *S* locus region between interspecific pairs has revealed that the *S* locus in *B. oleracea* is larger than that in *B. rapa*, the differences of the *S* locus sizes between them being due to insertion of retrotransposons. The retrotransposons identified in the *S*-locus region have been

Edited by Takashi Endo

* Corresponding author. E-mail: nishio@bios.tohoku.ac.jp

named *STF* (*S* locus retrotransposon family). Southern blot analyses using *STF* sequences as probes have revealed that *B. rapa* and *B. oleracea* have some *STF* homologs in common but that other *STF* homologs are present only in one species. More signals of *STF* homologs have been detected in *B. oleracea* than in *B. rapa* (Fujimoto et al., 2006a).

In the present study, we sequenced many homologs of five *STF*s, i.e., *BoSTF7a*, *BoSTF12a/15a*, *BoSTF12b*, and *BrSTF60a*, classified into a *Ty3-gypsy* group (Fujimoto et al., 2006a), and *BrSTFf2a,* classified into a *Ty1-copia* group (Fujimoto et al., 2006b). We also sequenced and phylogenetically analyzed *Tto1* in *Brassica*, classified into the *Ty1-copia* group.

MATERIALS AND METHODS

Plant Materials F_1 hybrid cultivars of Chinese cabbage, 'CR-Seiga 65' (Ishii Seed Co.) and 'Cream No. 2' (Watanabe Seed Co.); komatsuna, 'Osome' (Takii Seed Co.); pak-choi, 'Seibu' (Sakata Seed Co.); turnip, 'Wase-Ohkabu' (Takii Seed Co.); and oil seed rape, 'Yellow Sarson' (C634; Tohoku University Brassica Seed Bank, http://www.agri.tohoku.ac.jp/pbreed/Seed_Stock_DB/SeedStock-top.html) were used as materials of *Brassica rapa*. F_1 hybrid cultivars of broccoli, 'Greencomet' (Takii Seed Co.) and 'Ryokurei' (Sakata Seed Co.), cabbage, 'CM' (Takii Seed Co.), and cauliflower, 'Bridal' (Sakata Seed Co.), and inbred lines of kale and Chinese kale were used as materials of *Brassica oleracea*. Genomic DNA for Southern blot analysis was isolated from leaves by the CTAB method (Murray and Thompson, 1980).

Ten species in Brassicaceae, i.e., *B. carinata* (CA-114)*, B. juncea* (J-117)*, B. napus* ('Westar')*, B. nigra* (Ni-140), *B. tournefortii* (T-162), *Diplotaxis erucoides* (DIP-ERU-9)*, Eruca sativa* (ERU-SAT-1), *Raphanus sativus* (RAP-SAT-29)*, Sinapis alba* (SIN-ALB-25), and *Sinapis arvensis (*SIN-ARV-13), which are maintained in the Tohoku University Brassica Seed Bank, were used for analysis of retrotransposon sequences. Genomic DNA was isolated from single seeds according to Sakamoto et al. (2000). Retrotransposons were amplified by PCR using the primers listed in Table 1. Amplified DNA was cloned using pGEM-T Easy Vector System I (Promega, WI, USA).

Southern blot analysis Genomic DNA (2 μg) digested with *Eco*RI was electrophoresed on 1.0% agarose gel and transferred onto a nylon membrane (Nytran, Whatman, UK). The membrane was hybridized with a digoxigenin-labeled probe at 65°C. The *rvt* region of *BrSTF60a*, the *rnaseH* region of *BrTto1/BoTto1*, and the *gag* region of *BoSTF7a*, *BoSTF12a/15a*, *BoSTF12b*, and *BrSTFf2a* were used as probes. The *gag* region was mainly used as a probe because of its specificity. After hybridization, the membrane was washed twice in a solution of 0.1% SSC containing 0.1% SDS at 65°C for 20 min, and signals were detected according to an instruction manual (Roche, Rotkreuz, Switzerland).

Nucleotide sequence analysis Nucleotide sequences of 24 clones of PCR products were determined with a CEQ 2000XL DNA Analyzer (Beckman Coulter, CA, USA), and the data were analyzed using Sequencher (Gene Codes Corporation, MI, USA). The sequences were aligned using ClustalW (http://www.ddbj.nig.ac.jp/search/clustalw-j.html). The average number of substitution was calculated by PAUP ver. 4.0 (Swofford, 1998). Phylogenetic trees were constructed with the neighbor joining method (Saitou and Nei, 1987), and bootstrap probabilities of 1,000 trials were calculated.

RESULTS

The pattern of divergence of the retrotransposons was investigated using Brassicaceae species. The six putative retrotransposons (*i.e.*, putative homologs of five *STF*s and one *Tto1*) were isolated from *B. rapa* and *B. oleracea* genomic DNAs by PCR with the same primer sets and reaction cycle conditions. To reveal the copy number of *STFs* and *Tto1* in the *Brassica* genome, Southern blot analyses were performed using genomic DNAs of six cultivars each in *B. rapa* and *B. oleracea*. The number of bands detected by Southern blot analyses with six retrotransposon probes was evaluated using the following scores: 0, no band; 1, one band; 2, 2–5 bands; 3, more than 6 bands; and 4, smear (Fig. 1). The results for *B. rapa* and *B. oleracea* are summarized in Table 2. The copy number of retrotransposons in *B. oleracea* was generally larger than that in *B. rapa*. The average scores of all six

Table 1. Sequences of primers used for PCR amplifications

Gene	Length (bp)	Primer sequences (5'–3')	
		Forward	Reverse
BoSTF7a-gag	285	AAGTTATTCCCTTTCTCTTTAGGGG	AGGAAAAGCACCACGGTAGAATGTG
BoSTF12a/15a-gag	276	CATGAACTCCATTATAGAAAATG	TCCCGTTTTCCTTAGCTGATAAAGC
BoSTF12b-gag	281	CACCTGTTCGTCGAGAATCTAGAAG	CGTCGCCAGAGCGCCTTTCTGAGCG
BrSTF60a-rvt	498	CCGGAATGGCTCGCTAACCCAGTGG	CGGGGTATCGTCCGGAGAATTCCTT
BrSTFf2a-gag	264	ATGCACCTCACCGGTAAAGCCACGC	CATCTTCTTGGCAAACTTGAACCGC
Tto1-rnaseH	592	TGTATGTCGATGATATGTTGATTG	GCTCCACCCGCAAATGTTACCAAG

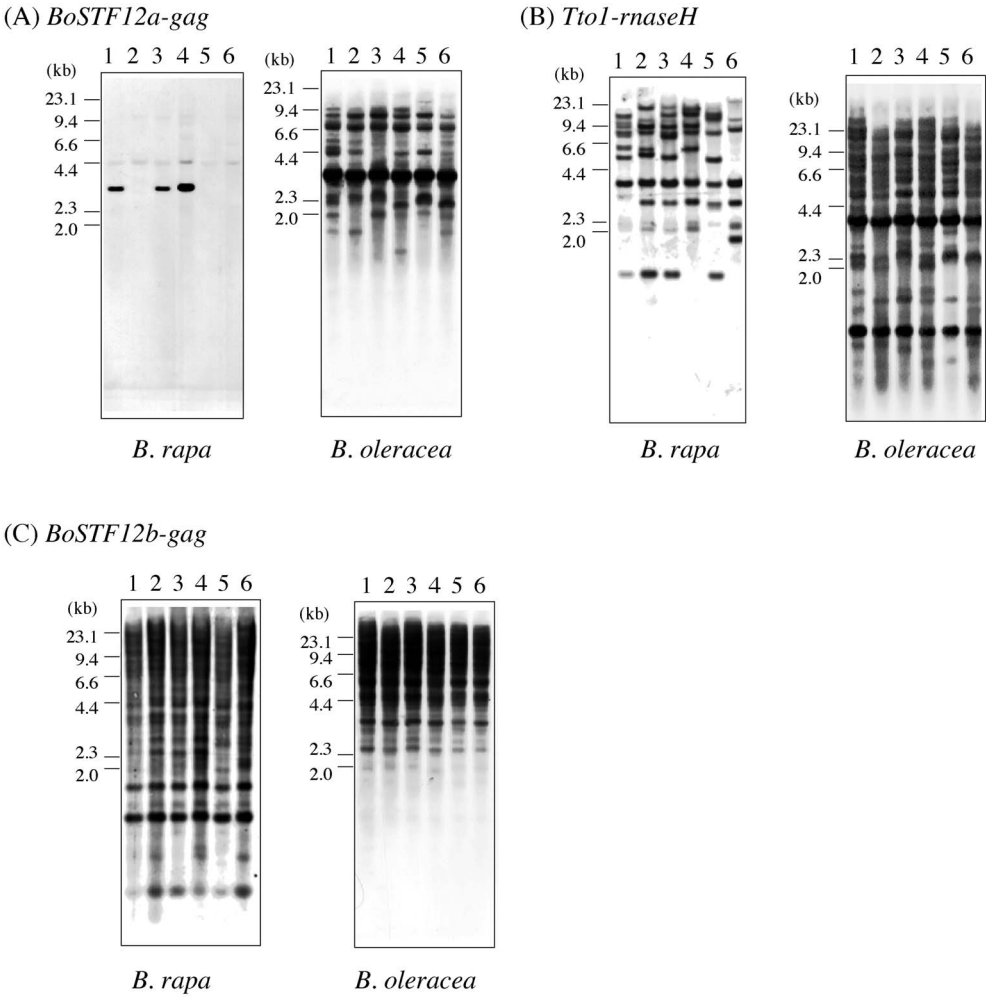

Fig. 1. Southern blot analysis of genomic DNAs of six cultivars in *B. rapa* and *B. oleracea*. After electrophoresis, genomic DNA digested with *Eco* RI was hybridized with the probes of the *gag* regions of *BoSTF12a* and *BoSTF12b* and the *rnaseH* region of *Tto1*. Lanes 1 to 6 in *B. rapa* are as follows: 1. 'Osome'; 2. 'CR-Seiga 65'; 3. 'Cream No. 2'; 4. 'Seibu'; 5. 'Wase-Ohkabu'; 6. 'Yellow Sarson'. Lanes 1 to 6 in *B. oleracea* are as follows: 1. 'Greencomet'; 2. 'Ryokurei'; 3. 'CM'; 4. 'Bridal'; 5. Chinese kale; 6. Kale. The scores estimated by Southern blot analyses are shown in Table 2.

Table 2. Scores of band numbers detected by Southern blot analyses using retrotransposon probes

	BoSTF7a-gag	*BoSTF12a/15a-gag*	*BoSTF12b-gag*	*BrSTF60a-rvt*	*BrSTFf2a-gag*	*Tto1-rnaseH*
B. rapa						
1. 'Osome'	4	1	4	2	1	3
2. 'CR-Seiga 65'	4	0	4	3	1	3
3. 'Cream No. 2'	4	1	4	3	1	3
4. 'Seibu'	4	1	4	3	1	3
5. 'Wase-Ohkabu'	4	0	4	3	0	3
6. 'Yellow Sarson'	4	0	4	2	2	3
Average	4.0	0.5	4.0	2.7	1.0	3.0
B. oleracea						
1. 'Greencomet'	4	3	4	4	2	4
2. 'Ryokurei'	4	3	4	4	1	4
3. 'CM'	4	3	4	4	0	4
4. 'Bridal'	4	3	4	4	0	4
5. Chinese kale	4	3	4	4	0	4
6. Kale	4	3	4	4	1	4
Average	4.0	3.0	4.0	4.0	0.7	4.0

0: no band; 1: one band; 2: 2–5 bands; 3: more than 6 bands; and 4: smear.

retrotransposons were 3.3 and 2.5 in *B. oleracea* and *B. rapa*, respectively.

Twenty-four randomly selected clones of PCR products for the six retrotransposons, which were amplified from genomic DNAs of *B. rapa* cv. 'Osome' and *B. oleracea* cv. 'Greencomet', were sequenced to reveal their divergence patterns in the genome. For each retrotransposon, phylogenetic relationships were inferred by the neighbor-joining method (Fig. 2). Species-specific clades of *B. rapa* and *B. oleracea* were observed in the phylogenetic trees of *BrSTF60a-rvt* and *BrSTFf2a*-gag. *Tto1-rnaseH* also showed species-specific clades except for one *B. rapa* sequence in the *B. oleracea* clade. The remaining sequences showed *trans*-specific patterns, indicating the retention of ancestral sequences of retrotransposons in various genomic regions.

The branch lengths in *B. oleracea* clades seem longer than those in *B. rapa*, implying higher diversity of the retrotransposons in *B. oleracea* (Fig. 2). Indeed, the average number of substitutions in pairwise comparison between the 24 retrotransposons revealed significantly more nucleotide substitutions in five of the six retrotransposons in *B. oleracea* than in *B. rapa* ($P < 0.01$; Table 3). The results of the average number of substitutions may reflect the difference of retrotransposon copy number between *B. oleracea* and *B. rapa*.

Since higher copy number and diversity of retrotransposons were found in *B. oleracea* than in *B. rapa*, whose genome size is about 100-Mb smaller than that of *B. oleracea* (Arumuganathan and Earle, 1991), we investigated the relationship between genome size and the number of substitutions in two retrotransposons using nine diploids of Brassicaceae. Twenty-four clones of homologs of *BoSTF7a* and *Tto1* amplified from the genomic DNAs

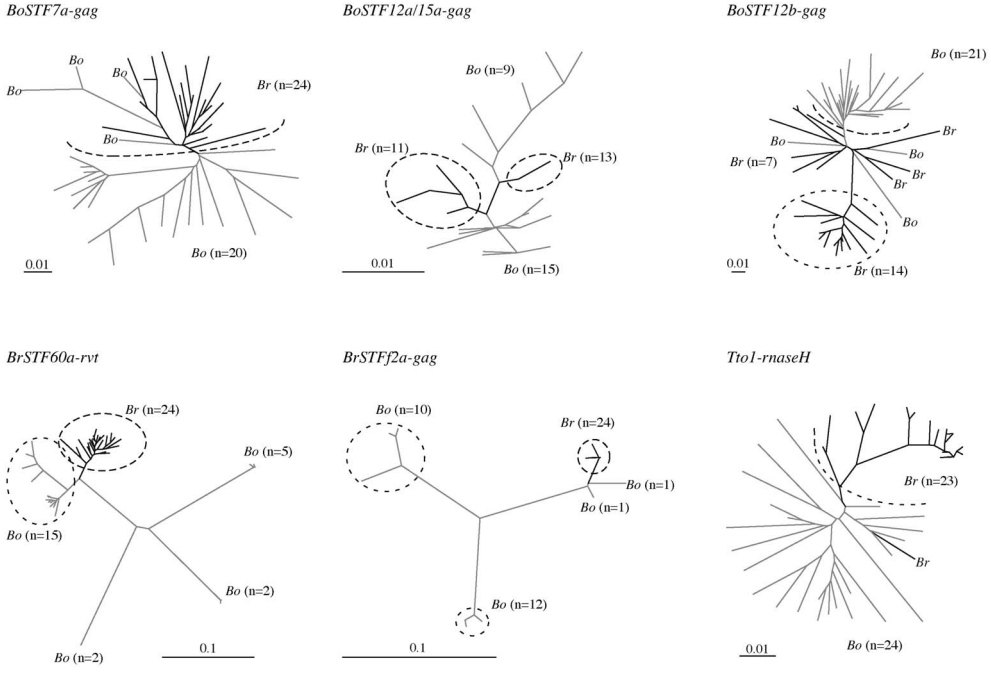

Fig. 2. Neighbor-joining trees of nucleotide sequences of the *gag* and *rvt* regions of *STF*s and the *rnaseH* region of *Tto1* in *B. rapa* and *B. oleracea*. Black and gray lines indicate the sequences of *B. oleracea* and *B. rapa*, respectively. Bo, *B. oleracea*; Br, *B. rapa.*

Table 3. Average number of substitutions of six retrotransposons in *B. rapa* and *B. oleracea*

	B. rapa	*B. oleracea*	
BoSTF7a-gag	0.0385 ± 0.00093	0.0691 ± 0.0013	$p < 0.01$
BoSTF12a/15a-gag	0.00952 ± 0.00048	0.0170 ± 0.00055	$p < 0.01$
BoSTF12b-gag	0.0981 ± 0.0026	0.0776 ± 0.0015	$p < 0.01$
BrSTF60a-rvt	0.0376 ± 0.00090	0.182 ± 0.0073	$p < 0.01$
BrSTFf2a-gag	0.00234 ± 0.00023	0.0982 ± 0.0045	$p < 0.01$
Tto1-rnaseH	0.0274 ± 0.0014	0.0489 ± 0.00082	$p < 0.01$

p: t-test of average number of substitutions between *B. rapa* and *B. oleracea*.

Table 4. Average number of substitutions of *BoSTF7a* homolog and *Tto1* homolog in Brassicaceae species

	Average number of substitutions		Genome size*
Species	*STF7a-gag*	*Tto1-rnaseH*	(Mbp/1C)
Brassica nigra	0.0744 ± 0.0025	0.0636 ± 0.00093	468
Brassica oleracea ssp. capitata	0.0691 ± 0.0013	0.0489 ± 0.00082	603
Brassica rapa ssp. chiensis	0.0385 ± 0.00093	0.0274 ± 0.0014	507
Brassica tournefortii	0.132 ± 0.0030	0.0542 ± 0.0010	791
Eruca sativa	0.0922 ± 0.0020	0.0442 ± 0.0010	560
Diplotaxis erucoides	0.0785 ± 0.0020	0.0387 ± 0.00093	632
Sinapis alba	0.0612 ± 0.0014	0.116 ± 0.0028	492
Sinapis arvensis	0.0634 ± 0.0020	0.0217 ± 0.00065	367
Raphanus sativus	0.0447 ± 0.0032	0.0486 ± 0.0020	526

*Arumuganathan and Earle (1991).

of nine diploids by PCR were sequenced. The estimated average number of substitutions and the genome size according to Arumuganathan and Earle (1991) are shown in Table 4. A significant positive correlation between genome size and the average number of substitutions of the *BoSTF7a* homologs in the nine species was observed (r = 0.726, P = 0.0261, permutation test) (Table 4, Fig. 3A), but genome size showed no correlation with the average number of substitutions in the *Tto1* homolog (r = 0.0465, P = 0.413, permutation test) (Table 4, Fig. 3B).

Based on cDNA sequences, the numbers of nonsense mutations and frameshift mutations in the 24 clones of the *BoSTF7a* homologs and the *Tto1* homologs were investigated in nine diploid species. There were more nonsense mutations than frameshift mutations in the *BoSTF7a* homologs. There were also more nonsense

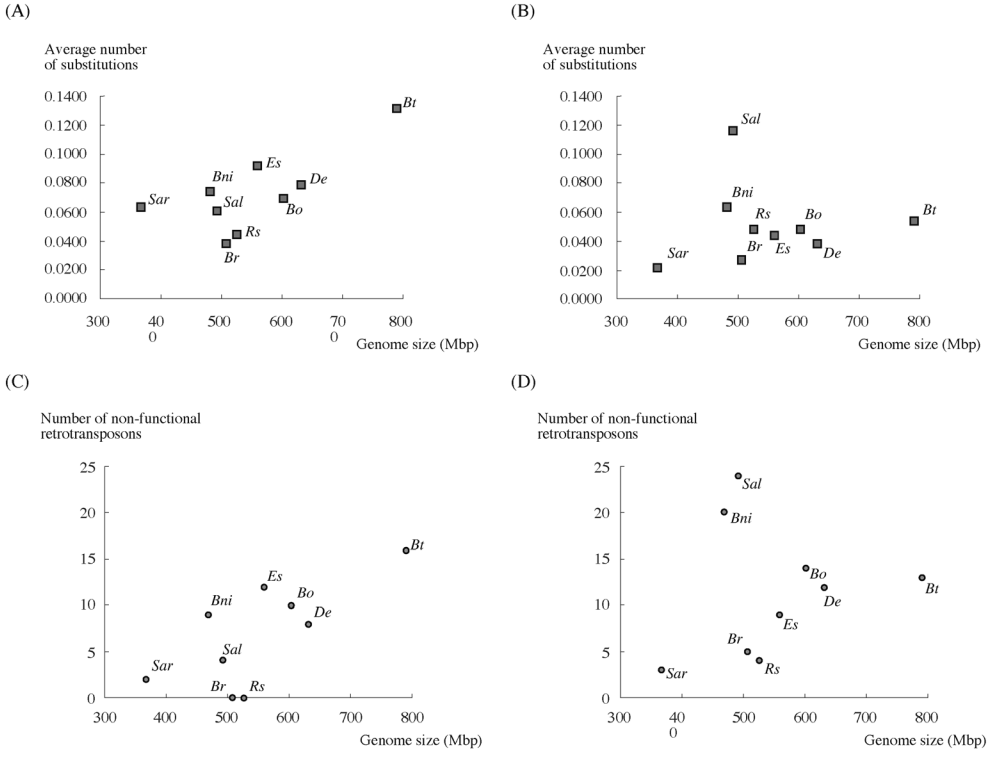

Fig. 3. Correlation between genome size and average number of substitutions in *STF7a* (A) and *Tto1* (B) (shown in Table 4) and between genome size and number of non-functional *STF7a* (C) and *Tto1* (D) (shown in Table 5) in nine diploids of Brassicaceae. The significance of the coefficient was computed by permutation test with 10,000 randomizations. Bni, *B. nigra*; Bo, *B. oleracea*; Br, *B. rapa*; Bt, *B. tournefortii*; De, *D. erucoides*; Es, *E. sativa*; Sal, *S. alba*; Sar, *S. arvensis*; Rs, *R. sativus*.

Table 5. Number of nonsense mutations and frameshift mutations for insertion and deletion of *BoSTF7a* homolog and *Tto1* homolog in Brassicaceae species

	BoSTF7a-gag			*Tto1-rnaseH*			
	Nonsense	Frameshift mutaion		Nonsense	Frameshift mutaion		Genome size*
Species	mutation	Insertion	Deletion	mutation	Insertion	Deletion	(Mbp/1C)
Brassica nigra	7	1(1)	1(1)	10	4(4)	6(18)	468
Brassica oleracea ssp. capitata	7	1(1)	2(3)	7	1(1)	6(54)	603
Brassica rapa ssp. chiensis	0	0	0	5	0	0	507
Brassica tournefortii	12	0	4(7)	5	1(8)	8(9)	791
Eruca sativa	6	1(3)	5(11)	2	3(6)	6(28)	560
Diplotaxis erucoides	4	1(3)	3(5)	3	0	11(40)	632
Sinapis alba	2	0	2(3)	5	1(1)	18(41)	492
Sinapis arvensis	2	0	0	1	0	2(4)	367
Raphanus sativus	0	0	0	1	3(16)	0	526

*Arumuganathan and Earle (1991).

Total lengths (bp) of insertion or deletion are shown in parentheses.

mutations and frameshift mutations by deletion than frameshift mutations by insertion in the *Tto1* homologs (Table 5). The number of deletions in the *BoSTF7a* sequences showed significant correlation with the genome size in nine diploids, but there was no correlation between the number of insertions in the *BoSTF7a* sequences and the genome size (Table 6). The numbers of deletions and insertions in *Tto1* showed no correlation with the genome size, either. Positive correlations between genome size and deletion size in the *BoSTF7a* sequences and between genome size and insertion size in the *Tto1* sequences were observed (Table 6). Genome size showed a correlation with the number of non-functional *BoSTF7a* sequences caused by nonsense mutations or frameshift mutations in the nine species ($r = 0.733$, $P = 0.0116$, permutation test) (Table 5, Fig. 3C), but showed no correlation with the number of non-functional *Tto1* homologs in those species ($r = 0.241$, $P = 0.272$, permutation test) (Table 5, Fig. 3D).

Table 6. Correlation coefficient between genome size and frequency/size of insertion and deletion in *STF7a* and *Tto1* homologs in Brassicaceae species

	BoSTF7a-gag		*Tto1-rnaseH*	
	r	*P*	*r*	*P*
Genome size/Number of insertion	0.146	0.348	–0.0648	0.455
Genome size/Number of deletion	0.671	0.0240	0.246	0.241
Genome size/Size of insertion	0.210	0.266	0.585	0.0603
Genome size/Size of deletion	0.586	0.0696	0.213	0.288

The divergence pattern of retrotransposons can provide further insight into the contribution of genome size evolution. Phylogenetic trees of nucleotide sequences of the *BoSTF7a* and *Tto1* homologs in nine diploids and three allopolyploids in Brassicaceae were inferred. Both the *BoSTF7a* and *Tto1* homolog sequences of eight diploid species clustered in one to three species-specific clades of retrotransposons (Figs. 4, 5). Multiple clades in each diploid species suggest that some types of retrotransposons may have been amplified and that others may have been lost after speciation. On the other hand, retrotransposons of *S. alba* and allopolyploids showed different patterns (Figs. 4, 5). The *BoSTF7a* homolog sequences, but not *Tto1*, of *S. alba* were clustered into clades of *B. rapa* and *B. oleracea*. The *BoSTF7a* and *Tto1* homolog sequences of allopolyploids were clustered into clades of their original genomes, *e.g.*, the sequences of *B. juncea* (AABB genome) belonging to the clade of *B. rapa* (AA genome) and that of *B. nigra* (BB genome).

DISCUSSION

Participation of retrotransposons in genome size variation More bands were generally detected in *B. oleracea* than in *B. rapa* by Southern blot analysis, indicating the presence of more retrotransposons in *B. oleracea* than in *B. rapa*, as was suggested in our previous study (Fujimoto et al., 2006a). The average number of substitutions of the five retrotransposons in *B. rapa* was also smaller than those in *B. oleracea*. The genome size of *B. oleracea* is larger than that of *B. rapa* (Arumuganathan and Earle, 1991). Therefore, we hypothesized that a species having a larger genome size in Brassicaceae has more copies of retrotransposons with larger nucleotide variation. Increase of copy number can simply result in more mutations, because mutations in retrotransposons may be selectively neutral or nearly neutral. A significant positive correlation was observed between genome size and average number of substitutions in *BoSTF7a* homologs in nine diploids, but no correlation between them was observed in the *Tto1* homologs. *B. nigra* and *S. alba* had exceptionally high percentages of non-functional *Tto1* homologs, more than 80% (Tables 5, 6), suggesting that *Tto1* in *B. nigra* and *S. alba* might have independently lost the activity of transposition after speciation. One possible scenario of accumulation of inactivated retrotransposons in *B. nigra* and *S. alba* is that, after speciation, both species might have independently experienced reduction of effective population size, such as a bottleneck event, leading to fixation of retrotransposons in the various genome positions by chance, which should have been followed by accumulation of null mutations. This scenario might also explain the exceptional pattern of *BoSTF7a* of *S. alba*, which did not form a species-specific clade, although alternative hypotheses such as horizontal gene transfer or introgression cannot be ruled out. If these two species are excluded as exceptions, there was a significant positive correlation between genome size and the average number of substitutions of *Tto1* homologs was obtained ($r = 0.821$, $P = 0.0012$, permutation test). These findings suggest that genome size of species, copy number of retrotransposons, and average number of substitutions in the retrotransposons correlate positively with each other.

Two retrotransposons may have contributed positively to the increase of genome size, in part, at least in the seven diploid species in Brassicaceae by amplification of their genome, supported by the phylogenetic patterns (Figs. 4, 5). Comparison of transposable elements between *A. thaliana* and *B. oleracea* has indicated that important factors involved in the genome size difference between *B. oleracea* and *A. thaliana* are proliferation of both class I and class II transposable elements and genome triplication in *Brassica* (Zhang and Wessler, 2004). There may be many transposons, such as the *STF7a* and *Tto1* homologs, which contribute to the genome size expansion in *Brassica*.

The transposition of retrotransposons is activated in plants under stress conditions and in a hypomethylated

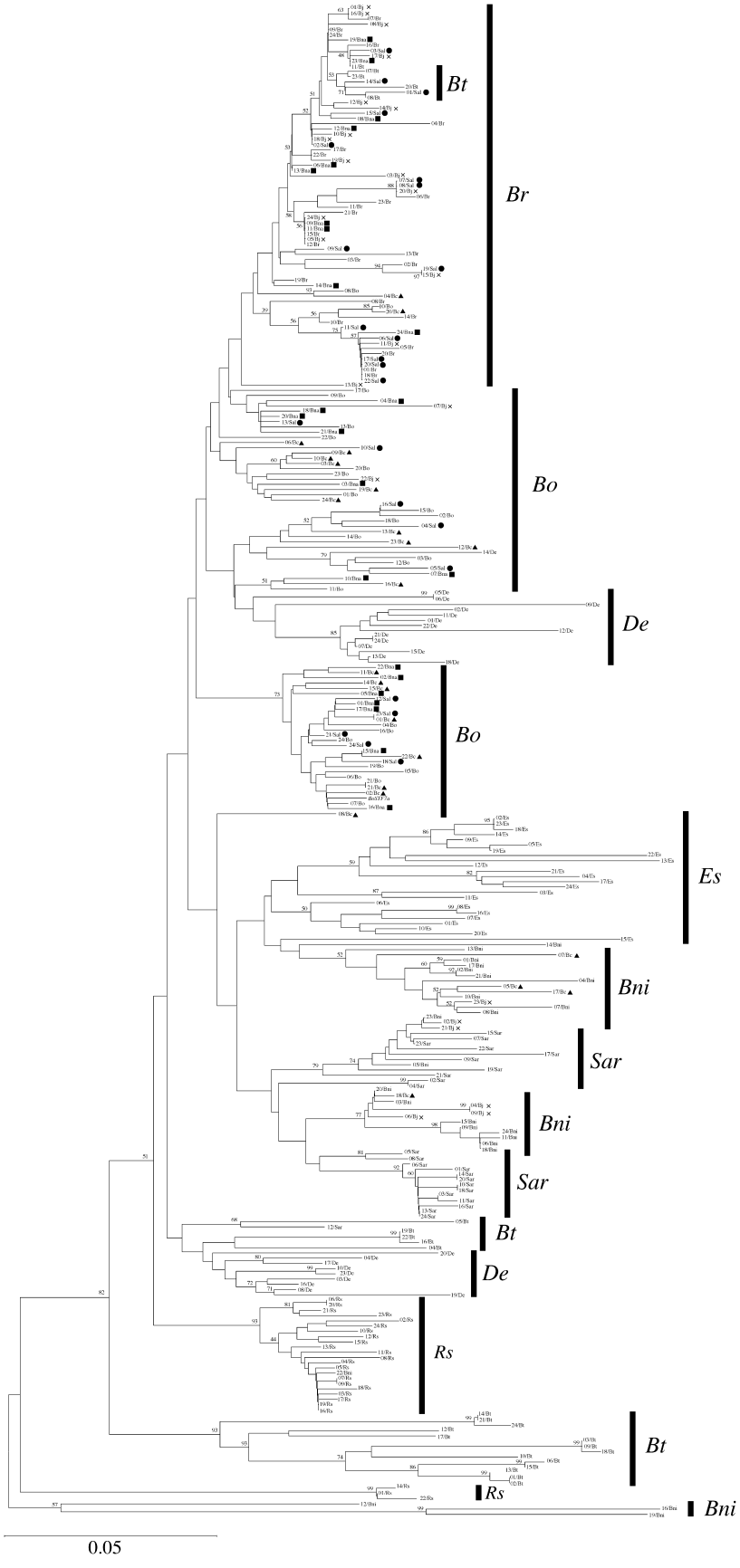

Fig. 4. A neighbor-joining tree of nucleotide sequences of the *gag* region of *STF7a*. Bootstrap values with 1,000 replicates are indicated at the node of the neighbor-joining trees. Bold crosses, filled triangles, squares, and circles represent the clones of allotetraploids species, *B. juncea*, *B. carinata*, *B. napus*, and diploid species *S. alba*, respectively. Bni, *B. nigra*; Bo, *B. oleracea*; Br, *B. rapa*; Bt, *B. tournefortii*; De, *D. erucoides*; Es, *E. sativa*; Sar, *S. arvensis*; Rs, *R. sativus*.

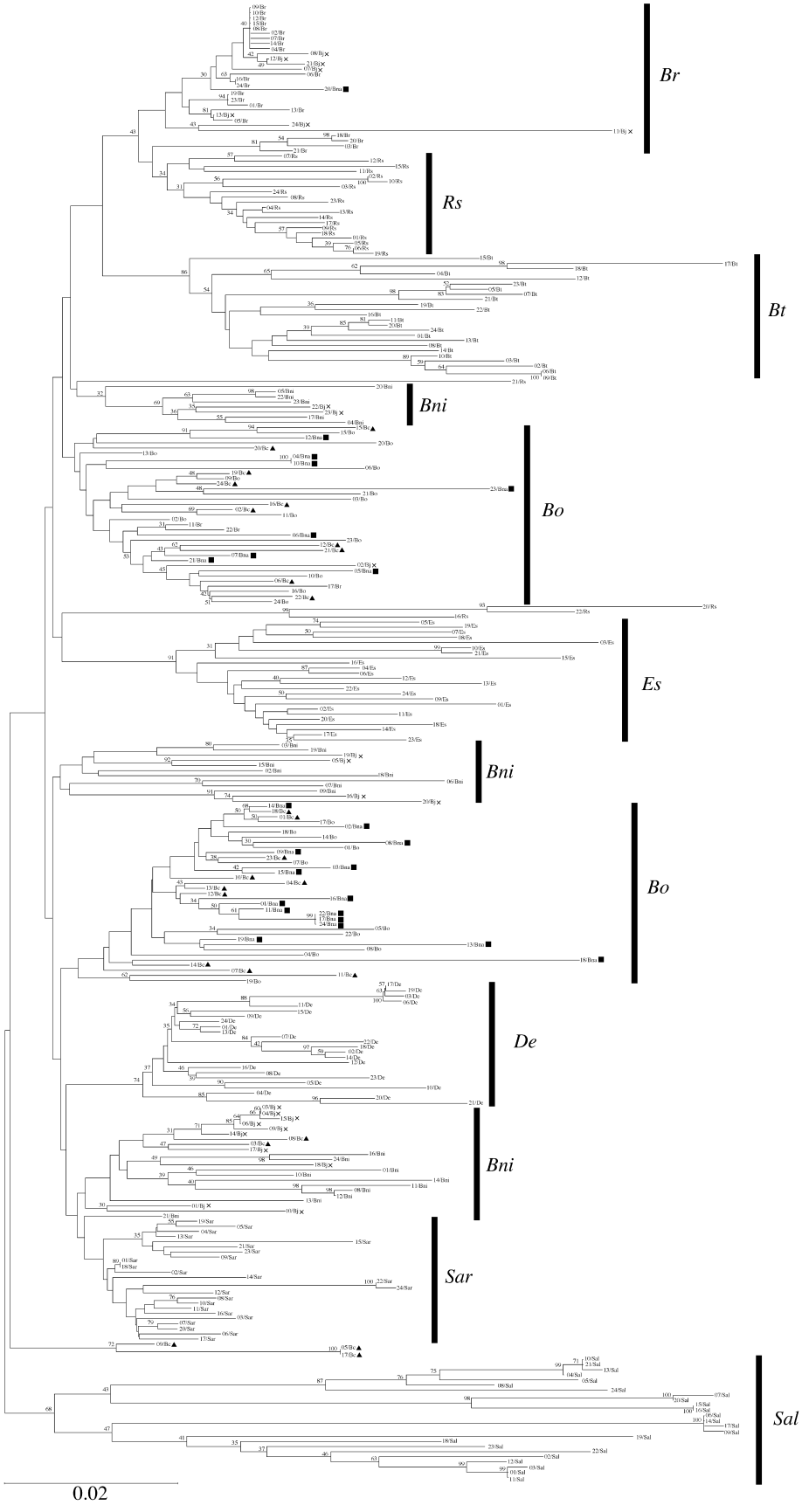

Fig. 5. A neighbor-joining tree of nucleotide sequences of the *rnaseH* region of *Tto1*. Bootstrap values with 1,000 replicates are indicated at the node of the neighbor-joining trees. Bold crosses, filled triangles, and squares represent the clones of allotetraploids species, *B. juncea*, *B. carinata*, and *B. napus*, respectively. Bni, *B. nigra*; Bo, *B. oleracea*; Br, *B. rapa*; Bt, *B. tournefortii*; De, *D. erucoides*; Es, *E. sativa*; Sal, *S. alba*; Sar, *S. arvensis*; Rs, *R. sativus*.

condition (Grandbastien, 1998; Hirochika et al., 2000). Such a situation might contribute to the difference of copy number of retrotransposons among the nine diploids in Brassicaceae, which would have experienced various environmental conditions. It is of great interest whether the difference of copy number of the *BoSTF7a* and *Tto1* homologs among the nine diploids was influenced by environmental stresses and epigenetic regulations. Since plant materials used in the present study consisted of only nine diploids in the tribe Brassiceae, further studies are needed to confirm the correlation between genome size and retrotransposon variations in different lineages.

Genome size and deletion of retrotransposons Another evolutionary mechanism of the change of genome size is deletion of retrotransposons leading to decreased size. Larger and more frequent deletions, more than 40 times, in a non-LTR retrotransposon, *Lau 1*, have been observed in *Drosophila melanogaster* than in the *Laupala* cricket, whose genome is 11 times larger than that of *D. melanogaster* (Petrov et al., 2000). More frequent and larger deletions during double-strand break repair can be seen in *Arabidopsis* than in tobacco, which has a twenty-fold larger genome than *A. thaliana* (Kirik et al., 2000). Our analysis, however, showed no negative correlation between genome size and frequency of short deletions in the *BoSTF7a* and *Tto1* homologs, but a positive correlation was observed between genome size and the number of deletions in the *BoSTF7a* homologs in the nine diploids. Reduction of genome size due to the loss of sequences was not observed in retrotransposons in Brassicaceae analyzed in the present study. However, *S. alba* and *B. nigra* had many inactivated retrotransposons and might be in the process of losing inactivated retrotransposons due to genetic drift.

Species-specific clustering of the *Ty1*-copia and *Ty3*-gypsy retrotransposon sequences The phylogenetic analysis of the present study showed that multiple species-specific clades were formed in the *BoSTF7a* homologs and the *Tto1* homologs, the first being classified into the *Ty3-gypsy* group and the second into the *Ty1-copia* group in the diploid species. As well as these two retrotransposons, phylogenetic analysis also showed that species-specific clades were also formed in the sequences of three other *Ty3-gypsy* and one *Ty1-copia* retrotransposons in *B. rapa* and *B. oleracea*. Multiple clades of each species suggest the maintenance of the ancestral variation in their genomes. In other words, in ancestral species, some types of sequences might have existed and two or three types might have been fixed in a species, while other types might have been lost in that species. All clades contained multiple similar sequences derived from a given species, indicating that one or a few types of retrotransposons may have been amplified after speciation. Thus, a high rate of amplification and loss of retrotransposons is required to explain the pattern of the phylogenetic tree. It has been suggested that the *Ty1-copia* retrotransposons diverged before the modern plant orders arose because of their high sequence heterogeneity (Konieczny et al., 1991; Flavell et al., 1992a, b; Voytas et al., 1992; VanderWeil et al., 1993; Noma et al., 1997). In genus *Vicia*, Hill et al. (2005) have suggested the possibility that different retrotransposon groups have different levels of activities. Low heterogeneity in both *Ty3-gypsy* and *Ty1-copia* retrotransposons observed in the present study suggest that the *Ty1*-copia retrotransposons as well as the *Ty3*-gypsy retrotransposons differentiated after speciation of the diploid species in Brassicaceae.

In the allopolyploid species in *Brassica*, no species-specific clades were formed, their retrotransposons were clustered in the clades of their original genomes, indicating that the amplification and differentiation of the *STF7a* homologs and the *Tto1* homologs took place before emergence of the allopolyploids. The nucleotide divergence of the two LTR sequences can be used for estimation of insertion time of LTR-retrotransposons (SanMiguel et al., 1998). Using the values of LTR sequence divergence, the insertion time of LTR-retrotransposons has been estimated to be within the last few million years in *A. thaliana*, barley, rice, tomato, and wheat (Bennetzen et al., 2005). The time of divergence between *B. nigra* and *B. oleracea* has been estimated to be a few million years ago (Yang et al., 1999). These estimations suggest that the amplification and differentiation of the *STF7a* and *Tto1* homologs might have occurred after speciation of various monogenomic species in Brassicaceae.

We thank Dr. S. Tsuchimoto for his helpful comments and suggestions. This work was supported in part by a grant-in-aid (19208001) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

Al-Shehbaz, I. A. (1984) The tribes of Cruciferae (Brassicaceae) in the southeastern United States. J. Arnold Arbor. **65**, 343–373.

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

Arumuganathan, K., and Earle, E. D. (1991) Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9**, 208–218.

Bateman, A. J. (1955) Self-incompatibility systems in angiosperms. III. Cruciferae. Heredity **9**, 52–68.

Bennetzen, J. L., Ma, J., and Devos, K. M. (2005) Mechanisms of recent genome size variation in flowering plants. Ann. Bot. **95**, 127–132.

Flavell, A. J., Dunbar, E., Anderson, R., Pearce, S. R., Hartley, R., and Kumar, A. (1992a) *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res. **20**, 3639–3644.

Flavell, A. J., Smith, D. B., and Kumar, A. (1992b) Extreme heterogeneity of *Ty1-copia* group retrotransposons in plants. Mol. Gen. Genet. **231**, 233–242.

Fujimoto, R., Okazaki, K., Fukai, E., Kusaba, M., and Nishio, T. (2006a) Comparison of the genome structure of the self-incompatibility (*S*) locus in interspecific pairs of *S* haplotypes. Genetics **173**, 1157–1167.

Fujimoto, R., Sugimura, T., Fukai, E., and Nishio, T. (2006b) Suppression of gene expression of a recessive *SP11/SCR* allele by an untranscribed *SP11/SCR* allele in *Brassica* self-incompatibility. Plant Mol. Biol. **61**, 577–587.

Grandbastien, M. A. (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci. **3**, 181–187.

Grandbastien, M. A., Spielmann, A., and Caboche, M. (1989) *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature **337**, 376–380.

Hill, P., Burford, D., Martin, D. M. A., and Flavell, A. J. (2005) Retrotransposon populations of *Vicia* species with varying genome size. Mol. Gen. Genomics **273**, 371–381.

Hirochika, H. (1993) Activation of tabacco retrotransposon during tissue culture. EMBO J. **12**, 2521–2528.

Hirochika, H., Sugimoto, K., Otsuki, Y., Tsugawa, H., and Kanda, M. (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc. Natl. Acad. Sci. USA **93**, 7783–7788.

Hirochika, H., Okamoto, H., and Kakutani, T. (2000) Silencing of retrotransposons in Arabidopsis and reactivation by the *ddm1* mutation. Plant Cell **12**, 357–368.

Johnston, J. S., Pepper, A. E., Hall, A. E., Chen, Z. J., Hodnett, G., Drabek, J., Lopez, R., and Price, H. J. (2005) Evolution of genome size in Brassicaceae. Ann. Bot. **95**, 229–235.

Kimura, R., Sato, K., Fujimoto, R., and Nishio, T. (2002) Recognition specificity of self-incompatibility maintained after the divergence of *Brassica oleracea* and *Brassica rapa*. Plant J. **29**, 215–223.

Kirik, A., Salomon, S., and Puchta, H. (2000) Species-specific double-strand break repair and genome evolution in plants. EMBO J. **19**, 5562–5566.

Konieczny, A., Voytas, D. F., Cummings, M. P., and Ausubel, F. M. (1991) A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics **127**, 801–809.

Kumar, A., and Bennetzen, J. L. (1999) Plant retrotransposons. Annu. Rev. Genet. **33**, 479–532.

Madsen, L. H., Fukai, E., Radutoiu, S., Yost, C. K., Sandal, N., Schauser, L., and Stougaard, J. (2005) LORE1, an active low-copy-number TY3-*gypsy* retrotransposon family in the model legume *Lotus japonicus*. Plant J. **44**, 372–381.

Murray, M. G., and Thompson, W. F. (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. **8**, 4321–4325.

Noma, K., Nakajima, R., Ohtsubo, H., and Ohtsubo, E. (1997) *RIRE1*, a retrotransposon from wild rice *Oryza australiensis*. Genes Genet. Syst. **72**, 131–140.

Nou, I. S., Watanabe, M., Isogai, A., and Hinata, K. (1993) Comparison of *S*-alleles and *S*-glycoproteins between two wild populations of *Brassica campestris* in Turkey and Japan. Sex. Plant Reprod. **6**, 79–86.

Ockendon, D. J. (2000) The *S*-allele collection of *Brassica oleracea*. Acta Hort. **539**, 25–30.

Petrov, D. A., Sangster, T. A., Johnston, J. S., Hartl, D. L., and Shaw, K. L. (2000) Evidence for DNA loss as a determinant of genome size. Science **287**, 1060–1062.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**, 406–425.

Sakamoto, K., Kusaba, M., and Nishio, T. (2000) Single-seed PCR-RFLP analysis for the identification of *S* haplotypes in commercial F_1 hybrid cultivars of broccoli and cabbage. Plant Cell Reports **19**, 400–406.

SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y., and Bennetzen, J. L. (1998) The paleontology of intergene retrotransposons of maize. Nature Genet. **20**, 43–45.

Sato, Y., Fujimoto, R., Toriyama, K., and Nishio, T. (2003) Commonality of self-recognition specificity of *S* haplotypes between *Brassica oleracea* and *Brassica rapa*. Plant Mol. Biol. **52**, 617–626.

Swofford, D. L. (1998) PAUP*. Phylogenetic analysis using parsimony (* and other methods), version 4.0 b. Sinauer Associates, Sunderland, Massachusetts.

U, N. (1935) Genome analysis in Brassica with special reference to the experimental formation of *B. napus* and peculiar mode of fertilization. Jpn. J. Bot. **7**, 389–452.

VanderWiel, P. L., Voytas, D. F., and Wendel, J. F. (1993) *Copia*-like retrotransposable element evolution in diploid and polyploidy cotton (*Gossypium* L.). J. Mol. Evol. **36**, 429–447.

Voytas, D. F., Cummings, M. P., Konieczny, A., Ausubel, F. M., and Rodermel, S. R. (1992) *copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA **89**, 7124–7128.

Xiong, Y., and Eickbush, T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9**, 3353–3362.

Yang, Y. W., Lai, K. N., Tai, P. Y., and Li, W. H. (1999) Rates of nucleotide substitution in angiosperm mitochondrial DNA sequences and dates of divergence between *Brassica* and other angiosperm lineages. J. Mol. Evol. **48**, 597–604.

Zhang, X., and Wessler, S. R. (2004) Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. Proc. Natl. Acad. Sci. USA **101**, 5589–5594.