Miniature Transposable Elements (mTEs): Impacts and Uses in the *Brassica* Genome

6

Perumal Sampath, Jonghoon Lee, Feng Cheng, Xiaowu Wang and Tae-Jin Yang

Abstract

Transposable elements occupy large portions of eukaryotic genomes and play an important role in genome evolution. Terminal repeat retrotransposons in miniature (TRIMs), short interspersed elements (SINEs) and miniature inverted-repeat transposable elements (MITEs) are representative forms of so-called miniature transposable elements (mTEs), which are present in very high-copy numbers, stable, widely distributed and in close association with genic regions in plant genomes. These features make mTEs useful for applications such as developing marker systems, functional characterization of associated genes, and elucidating the contribution of TEs to gene evolution. Here, we summarize the characteristics, copy numbers and distribution patterns of five TRIM families, 14 short interspersed elements (SINE) families and 20 MITE families in the *Brassica rapa* genome. We also show the comparative distribution pattern of paralogous mTE family members in *Brassica oleracea* and 11 *B. rapa* accessions. In addition, we describe putative roles for mTEs in the evolution of the triplicated *Brassica* genome and discuss the utility of mTEs for analysis of genome evolution and for developing practical marker systems.

6.1 Introduction

Transposable elements, also known as "mobile genetic elements", are DNA sequence fragments that move or are copied from one location to another in the genome either directly, by a cut-and-paste mechanism (class II DNA transposons), or indirectly, by a copy-and-paste mechanism through an RNA intermediate (class

P. Sampath · J. Lee · T.-J. Yang (✉)
Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea
e-mail: tjyang@snu.ac.kr

F. Cheng · X. Wang
Chinese Academy of Agricultural Sciences, Institute of Vegetables and Flowers, Zhongguancun Southern Street 12, 100081 Beijing, China

X. Wang and C. Kole (eds.), *The Brassica rapa Genome*, Compendium of Plant Genomes,
DOI 10.1007/978-3-662-47901-8_6

I retrotransposons; Wicker et al. 2007; Feschotte et al. 2002) (Fig. 6.1). Transposition of both classes of elements may result in a heritable increase in copy number within the genome; hence, individual TE types are found in multiple copies (often referred to as a TE family) and constitute the majority of the repetitive fraction of eukaryotic genomes. The large-scale sequencing of eukaryotic genomes has revealed that TEs are the most abundant component of most eukaryotic genomes, are present ubiquitously, and occupy large fractions of genomes: TEs account for 40 % of *Oryza sativa* (rice) (Feschotte 2008), 50 % of *Glycine max* (soybean) (Schmutz et al. 2010), and >80 % of *Zea mays* (maize), *Triticum aestivum* (wheat) and *Hordeum vulgare* (barley) (Paterson et al. 2009; Wicker et al. 2009; Bennett and Smith 1976) genomes. Whole-genome analyses estimated that ~40 % of the *Brassica rapa* ($2n = 2x = 529$ Mb) and *B. oleracea* ($2n = 2x = 696$ Mb) genomes are occupied by transposon-related sequences (Wang et al. 2011; Liu et al. 2014). High proportions of TEs are intact in *B. rapa* and *B. oleracea* genomes (68 and 98 %, respectively), although TEs have been continuously amplified in both genomes since at least 4.6 million years ago (MYA) (Liu et al. 2014). Compared to *B. rapa*, *B. oleracea* has many younger TEs, which are responsible for its increased genome size (Liu et al. 2014). Amplification of TEs in the genome can not only cause an increase in genome size but also help to drive the evolution of genes and genomes (Feschotte et al. 2002; Bire and Rouleux-Bonnin 2012; Feschotte 2008; Alzohairy et al. 2013), although most TEs are inactive

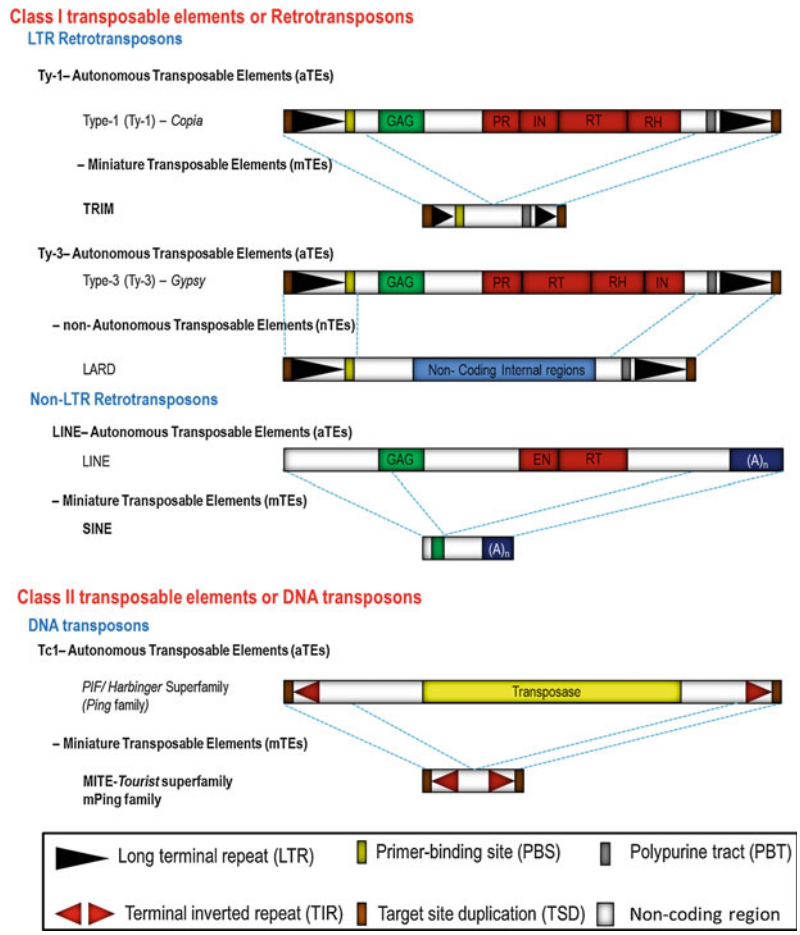

Fig. 6.1 Classification of TEs and mTEs. *LARD* large retrotransposon derivative; *TRIM* terminal-repeat retrotransposons in miniature; *LINE* long interspersed nuclear element; *SINE* short interspersed nuclear element; *GAG* a structural protein for virus-like particles; *PR* protease; *IN* integrase; *RT* reverse transcriptase; *RH* RNAse H; *EN* endonuclease

and mainly controlled by epigenetic mechanisms (e.g., DNA and histone methylation) (Hollister and Gaut 2009; Lisch 2009, 2012; Casacuberta and Santiago 2003).

TEs containing their own functional genes for transposition are referred to as autonomous transposable elements (aTEs), whereas TEs that lack coding genes and therefore cannot produce their own transposase or reverse transcriptase are termed nonautonomous or noncoding transposable elements (nTEs) (Casacuberta and Santiago 2003). The nTEs, such as large retrotransposon derivatives (LARDs), terminal repeat retrotransposons in miniature (TRIMs), short interspersed elements (SINEs) and miniature inverted-repeat transposable elements (MITEs), are generally deletion derivatives of aTEs and require a *trans*-acting transposase from their corresponding autonomous partner elements for transposition. TRIMs, SINEs and MITEs are examples of miniature transposable elements (mTEs; Fig. 6.1) (Casacuberta and Santiago 2003; Feschotte and Pritham 2007; Wessler et al. 1995; Okada et al. 1997), and families belonging to each type of mTE have had significant influences on gene and genome evolution (Wessler 2006; Witte et al. 2001). In this chapter, we summarize the characteristics, copy numbers and comparative distribution of 39 mTE families in *B. rapa* and *B. oleracea*. We go on to discuss the utility of mTEs for genomics-assisted breeding and evolutionary studies.

6.2 Characteristics and Distribution of mTEs

mTEs have unique structural characteristics and are ubiquitously present in eukaryotic genomes. The important characteristics of mTEs are summarized in Table 6.1.

Table 6.1 Characteristics of mTEs

Characteristics	TRIM	SINE	MITE
Element size	~900 bp (up to 2.5 Kb)	<700 bp	~800 bp (up to 2 Kb)
Structure			
Terminal repeats (bp)	TDR (100–350 bp)	No	TIR (10–1000 bp)
Target site duplication (TSD)	5 bp	No	2–11 bp (7 families identified)
Origin	Type-1 Copia LTR-RT	LINE	DNA-TE
Copy number	Moderately high	High	Very high
Copies in rice genome	~350	~2500	22,000–36,000
Copies in *Brassica* genome	~2000	~5000	~48,000
First identified element First identified organism (copies)	*Katydid* *Solanum tuberosum* (540)	*TS* *Nicotiana tabacum* (5400)	*Zm-1* *Zea mays* (1000–10,000)
Applications	Insertion polymorphism markers TRIM display	Insertion polymorphism markers SINE display	Insertion polymorphism markers MITE display
Discovery tools			
Structure based	LTR_Finder, LTR_STRUC, LTR_MINER, TRANSPO	SINEDR	MITEHunter, MITE Digger
Homology based	HMMER, repbase, repeat masker	SINEBase, repbase, repeat masker	P-MITE, repbase, repeat masker

6.2.1 TRIMs

TRIM elements are nonautonomous terminal direct repeat (TDR) retrotransposons with similar, but smaller, structural characteristics to long terminal repeat (LTR) retrotransposons. The TDR ranges from 100 to 350 bp, and has a 5-bp target site duplication (TSD). The internal region of 150–500 bp begins with a tRNA-methionine primer binding site (PBS) and ends with a poly-purine tract (PPT) motif (Witte et al. 2001; Yang et al. 2007; Fig. 6.1). It has been suggested that TRIM elements are mobilized through copy-and-paste mechanisms and that the proliferation of TRIMs has occurred with the help of *trans*-acting autonomous partner elements like Type-1/COPIAs, with which TRIMs share the characteristic signature structure, but there are no reports confirming this as yet (Witte et al. 2001; Kalendar et al. 2008).

TRIM elements are abundant and widely distributed and have been identified in both monocot and dicot plants and rarely in animals. *Katydid* was the first TRIM identified, originally in *Solanum tuberosum* (potato) and subsequently in the *Arabidopsis* genome (Witte et al. 2001). *Cassandra*, a unique TRIM family harboring 5S rDNA sequence, has been identified in more than 50 plant species including ferns and trees and is thought to have evolved 250 MYA (Kalendar et al. 2008; Sampath and Yang 2014a, b). Among the five TRIM families found in *Brassica* (TB-1-5), a total of 1393 and 1639 members were identified 283 Mb *B. rapa* sequences (including 256 Mb pseudo-chromosome and 27 Mb unanchored scaffold sequences) and 385 Mb *B. oleracea* pseudo-chromosome sequences. TRIM families are distributed throughout the *B. rapa* and *B. oleracea* genomes (Murukarthick et al. 2014). Though the insertions of five *Brassica* TRIM families were random, Genome-wide characterization showed that 619 (44 %) and 656 (40 %) of the members of the five *Brassica* TRIM families reside in or within 2 kb of a gene in the *B. rapa* and *B. oleracea* genome, respectively (Yang et al. 2007; Murukarthick et al. 2014; Table 6.2). Most-TRIM families are present in relatively similar copy numbers between *Brassica* species but the *Cassandra* family appears to show high divergence between *B. rapa* and *B. oleracea*, based on members found in counterpart paralogous sequences of pseudo-chromosomes (Sampath and Yang 2014a, b).

6.2.2 SINEs

SINEs are relatively short (75–662 bp), nonautonomous, non-LTR retrotransposons. SINEs have a unique structure composed of a head, body and tail. The head (5' end), which consists of an internal promoter, is derived from cellular RNA usually tRNA, 7SL RNA and/or 5S rRNA. SINEs lack a TSD but contain monopolymer tails. The internal promoter provides a transcription signal for transcription of the SINE by RNA polymerase III. The body of the SINEs originates either from autonomous partner elements or from distant SINE families. The 3' end (tail) of SINEs consists mostly of simple repeats. It has been suggested that SINEs also need a *trans*-acting partner (most probably LINEs) for amplification and mobilization (Okada et al. 1997; Kramerov and Vassetzky 2011a, b; Kramerov and Vassetzky 2005) (Fig. 6.1).

SINEs are abundant, present in high copy number and occupy a significant fraction of eukaryotic genomes. For instance, Alu, a well-characterized SINE from primates, exists in >1,500,000 copies in the human genome, covers >11 % of the total genome and played an important role in human population genetics and evolution (Venter et al. 2001; Batzer and Deininger 2002). In the *Brassicaceae*, 16 SINE families have been reported, including 1270 and 2364 members in the *B. rapa* and *B. oleracea* genomes, respectively (Vassetzky and Kramerov 2013). SINEs are distributed in various genomic locations of *B. rapa* and *B. oleracea*, with 599 (47.1 %) and 1154 (48.8 %) of the members, respectively, present in close association with genic regions with <2 kb of a gene (Murukarthick et al. 2014).

Table 6.2 Distribution of the members of 39 mTE families in the *B. rapa* and *B. oleracea* pseudo-chromosome sequences and 1× WGS data

mTE no.	mTE type	mTE ID[a]	Unit size (bp)	*B. rapa*			*B. oleracea*		
				Copies in genome assembly[b]	Estimated copies based on 1× WGS[c]	Range (copies)[c]	Copies in genome assembly[d]	Estimated copies based on 1× WGS[e]	Range (copies)[e]
1	TRIM	TB-1	364	72	120 ± 14.6	95–156	69	147 ± 15.5	131–162
2	TRIM	TB-2	387	21	61 ± 10.7	45–81	4	105 ± 2.5	102–107
3	TRIM	TB-3	1313	1	10 ± 1.4	7 ± 12	1	21 ± 1.5	19–22
4	TRIM	TB-4	598	43	132 ± 6.6	122–143	55	276 ± 15.5	260–291
5	TRIM	TB-5	781	19	128 ± 36.5	71–208	131	190 ± 39.5	150–229
6	SINE	SB-1	171	0	0 ± 0.3	0–1	14	50 ± 2	48–52
7	SINE	SB-2	149	0	23 ± 3.8	15–30	0	33 ± 3	30–36
8	SINE	SB-3	297	19	30 ± 7.6	17–47	86	184 ± 2	182–186
9	SINE	SB-5	162	0	51 ± 19.2	28–89	112	196 ± 7.5	188–203
10	SINE	SB-6	297	55	58 ± 15.6	25–81	93	279 ± 23.5	255–302
11	SINE	SB-7	352	16	63 ± 11.5	39–81	43	155 ± 11	144–166
12	SINE	SB-8	95	0	0 ± 0	0–0	64	57 ± 0.5	56–57
13	SINE	SB-9	212	80	83 ± 27.1	56–154	101	120 ± 11	109–131
14	SINE	SB-10	159	5	5 ± 1.9	2–9	76	70 ± 10	60–80
15	SINE	SB-11	170	0	1 ± 0.5	1–2	32	38 ± 1	37–39
16	SINE	SB-12	170	3	5 ± 1.1	3–7	41	46 ± 6	40–52
17	SINE	SB-13	225	0	0 ± 0.5	0–1	12	40 ± 14.5	25–54
18	SINE	SB-14	156	11	31 ± 10.4	18–54	50	74 ± 0.5	73–74
19	SINE	SB-15	206	0	1 ± 0.4	0–2	1	0 ± 0	0–0
20	MITE	BraSto-1	267	16	32 ± 22.8	5–92	50	249 ± 44	205–293
21	MITE	BraSto-2	260	401	155 ± 97.2	32–392	210	671 ± 137	534–808
22	MITE	BraSto-3	242	6	2 ± 1.3	0–5	2	2 ± 1	1–3
23	MITE	BraSto-4	558	97	38 ± 35.7	6–138	336	489 ±146.5	342–635
24	MITE	BraTo-2	212	8	44 ± 22.5	11–92	127	986 ± 135.5	850–1121
25	MITE	BraTo-1	366	61	25 ± 9.8	9–40	60	117 ± 25.5	91–142
26	MITE	BraTo-3	252	245	152 ± 70.4	43–291	116	212 ± 22	190–234
27	MITE	BraTo-4	160	287	217 ± 66.2	88–335	36	91 ± 10.5	80–101
28	MITE	BraTo-5	286	118	52 ± 28.6	14–120	37	93 ± 15.5	77–108
29	MITE	BraTo-6	257	60	14 ± 5.3	5–24	76	142 ± 35.5	106–177
30	MITE	BraTo-7	366	54	25 ± 14.3	4–53	199	471 ± 95.5	375–566
31	MITE	BraTo-8	348	29	14 ± 9	2–35	26	66 ± 26	40–92
32	MITE	BraTo-9	264	20	72 ± 24.4	31–127	32	30 ± 9.5	20–39
33	MITE	BraTo-10	255	35	28 ± 14.2	8–58	50	105 ± 19.5	85–124
34	MITE	BraTo-11	305	4	2 ± 1.5	0–6	5	5 ± 2	3–7
35	MITE	BraTo-12	273	66	31 ± 17	7–62	67	81 ± 7	74–88

(continued)

Table 6.2 (continued)

mTE no.	mTE type	mTE ID[a]	Unit size (bp)	*B. rapa*			*B. oleracea*		
				Copies in genome assembly[b]	Estimated copies based on 1× WGS[c]	Range (copies)[c]	Copies in genome assembly[d]	Estimated copies based on 1× WGS[e]	Range (copies)[e]
36	MITE	BraTo-13	268	74	35 ± 15.4	11–62	85	177 ± 26	151–203
37	MITE	BraHAT-1	439	24	15 ± 7.7	5–30	55	193 ± 37.5	155–230
38	MITE	BraHAT-2	248	16	23 ± 12.9	6–58	19	23 ± 1	22–24
39	MITE	BraMu-1	271	24	12 ± 7.2	2–28	16	87 ± 21	66–108
			Total	1990	1790		2589	6371	

[a]SB-4 and SB-16 are not shown here because they are not present in either *B. rapa* or *B. oleracea* but are present in *A. thaliana*

[b]mTE copies were identified from the available 283 Mb whole-genome pseudo-chromosome sequences of *B. rapa* with 80 % sequence similarity

[c]Mean, standard deviation (SD) and range were calculated based on members estimated from 1× (529 Mb) WGS of 11 *B. rapa* accessions

[d]mTE copies were identified from the available 385 Mb whole-genome pseudo-chromosome sequences of *B. oleracea* with 80 % sequence similarity

[e]Mean, standard deviation (SD) and range were calculated based on members estimated from 1× (696 Mb) WGS of 2 *B. oleracea* accessions

6.2.3 MITEs

MITEs are class II nonautonomous TEs characterized by relatively small size (<800 bp), AT-rich sequences, and flanking terminal inverted repeats (TIRs) ranging from 10 to 200 bp (Sampath and Yang 2014a, b). Insertion of a MITE can produce TSD ranging from 2 to 11 bp depending on the MITE superfamily involved (Fig. 6.1) (Bureau and Wessler 1994; Lu et al. 2012). The TIRs are more conserved than their respective internal sequences, and act as a recognition site for endonucleases for integration of TEs via transposition (Casacuberta and Santiago 2003). The TIRs are complementary to each other, leading to the formation of a secondary loop structure, which can be a source of small RNA and may act in gene regulation (Mo et al. 2012; Sampath et al. 2013; Sarilar et al. 2011). The internal sequences of MITEs have sequence diversity due to the influence of unrelated autonomous TEs during transposition (Sampath et al. 2013; Yaakov and Kashkush 2012). Unlike TRIMs and SINEs, transposition of MITEs occurs by cut-and-paste mechanisms, and MITEs can be amplified in the genome by abortive gap repair, bursts of amplification, or as yet unknown mechanisms under stress (Fattash et al. 2013; Casacuberta 2013).

Different MITE families are classified based upon TSD length, structure, and sequence similarity to the putative transposase of the corresponding DNA transposon. MITEs were first identified in the maize genome and later in various other plant and animal genomes (Bureau and Wessler 1994, 1992; Feschotte and Wessler 2002). So far, seven MITE superfamilies have been identified in plants, although 15 superfamilies of DNA transposons have been reported (Fattash et al. 2013). MITEs comprise two major families, namely *Stowaway*-like (with TA as the TSD) and *Tourist*-like (with TAA as the TSD), as well as several other minor families including

Table 6.3 Plant materials used for re-sequencing and mTE analysis

	ID	Morphotype	Species	Sub species	Accession (cultivar)	Genome	WGS reads for mTE analysis	
							Amounts (Mbp)	Coverage (x)
1	Br-1	Chinese cabbage	*B. rapa*	ssp. *pekinensis*	Chiifu	AA	2321.4	4.4
2	Br-2	Chinese cabbage	*B. rapa*	ssp. *pekinensis*	Kenshin	AA	1498.9	2.8
3	Br-3	Chinese cabbage	*B. rapa*	ssp. *pekinensis*	DF10C062	AA	1410.9	2.7
4	Br-4	Chinese cabbage	*B. rapa*	ssp. *pekinensis*	Z16	AA	1496.2	2.8
5	Br-5	Turnip Asian	*B. rapa*	ssp. *rapifera*	Yoya	AA	1495.9	2.8
6	Br-6	Rapini-Caixin	*B. rapa*	ssp. *parachinensis*	L58	AA	1492.5	2.8
7	Br-7	Pak Choi	*B. rapa*	ssp. *chinensis*	Suzhouqing	AA	1495.3	2.8
8	Br-8	Canola	*B. rapa*	ssp. *oleifera*	R-o-18	AA	1497.8	2.8
9	Br-9	Mizuna	*B. rapa*	ssp. *nipposinica*	Mizuna	AA	1497.5	2.8
10	Br-10	Turnip Europe	*B. rapa*	ssp. *rapifera*	Manchester	AA	1484.5	2.8
11	Br-11	Canola-rapid cycling	*B. rapa*	ssp. *oleifera*	L144	AA	1496.6	2.8
12	Bo-1	Cabbage	*B. oleracea*	ssp. *capitata*	C1176	CC	1541	2.2
13	Bo-2	Cabbage	*B. oleracea*	ssp. *capitata*	C1220	CC	1606.8	2.3

1–11 WGS of *B. rapa* accessions were kindly provided by Xiaowu Wang (Key Laboratory of Horticultural Crops Genetic Improvement of Ministry of Agriculture, Sino-Dutch Joint Lab of Horticultural Genomics Technology, Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing, China). *12–13* WGS of *B. oleracea* was generated with support of a grant from the *Golden Seed Project (Center for Horticultural Seed Developmenet,* ***No. 213003-04-3-SB430****), Ministry of Agriculture, Food and Rural Affairs(MAFRA)*

hAT-like (with 5, 6, or 8 bp TSDs), *MULE* (with 9–10 bp TSDs), and *En/Spm* (3-bp TSDs) MITEs (Oki et al. 2008).

MITEs include the mTE members with the most copies, distributed throughout the genome. MITE family members occupy different proportions in plant and animal genomes, reaching up to 10 % in rice, 8 % in *Medicago*, 4 % in *B. rapa,* 0.71 % in *A. thaliana* and 16 % in *Aedes aegypti* (yellow fever mosquito) (Paterson et al. 2009; Schmutz et al. 2010; Lu et al. 2012; Nene et al. 2007; Chen et al. 2013). *In silico* analysis reveals 174 families with more than 45,821 members including *Tourist*, *Stowaway* (Feschotte and Pritham 2007), *Mutator* (Alzohairy et al. 2013) and *CACTA* (Wicker et al. 2007) in the *B. rapa* genome. Furthermore, 20 MITE families including two novel families were identified in *B. rapa* and *B. oleracea*, and comparative analysis providing useful information for genomics and breeding (Chen et al. 2013; Sampath et al. 2014).

6.3 Identification of mTEs

There are various bioinformatics tools available for mining of mTEs within genomes, each with its own advantages and drawbacks (Janicki et al. 2011). Sequence similarity-based analysis tools require a known repeat library for sequence searches, whereas structure-based mTE mining tools promote identification of novel families but can have false positive rates of up to 86 % (Han and Wessler 2010). Currently, 40 different mTE families, including five TRIM, 16 SINE and 20 MITE families, and their member distribution in

B. rapa and *B. oleracea* are listed in a recently developed database, BrassicaTED (http://im-crop.snu.ac.kr/BrassicaTED/index.php). BrassicaTED also includes tools for mining and characterization of mTEs and TEs (Murukarthick et al. 2014).

Genomic tools currently available for genome-wide surveys of TRIM elements include LTR_finder, LTR_STRUC, LTR_MINER, TRANSPO, and find_ltr, which use structure-based approaches, as well as Repbase and Repeatmasker based on sequence similarity.

For SINE families, comprehensive, up-to-date information about structural characteristics and copy numbers can be found in the SINEBase database (Vassetzky and Kramerov 2013). SINEBase, Repbase and Repeatmasker can be used for homology searches of SINE families. Most novel SINE families have been identified using SINE consensus sequences, such as the tRNA-related portions. The SINEDR tool has also been used to identify novel SINE families.

Mining of MITEs on the genome scale can be done using various genomics tools. For instance, FINDMITE (Tu 2001), MUST (Chen et al. 2009), MITE Hunter (Han and Wessler 2010), RSBP (Lu et al. 2012) and MITE digger (Yang 2013) are available online to identify MITEs based on signature structures such as the TIR and TSD. Repbase, Repeatmasker, Inverted Repeat Finder, REPuter, RECON, Micropeats and STAN can also be used to mine the MITEs based on sequence similarity (Smit and Hubley 1996; Warburton et al. 2004; Jurka 2000; Lerat 2009). A recently developed database for plant MITEs (P-MITE) contains MITEs from 40 different species including *Brassica*.

6.4 Influence of mTEs on Evolution of the Triplicated *Brassica* Genome

The mTEs can play roles in remodeling gene structures by exon shuffling and in gene expression modification by providing new transcription start sites, splicing sites and poly-A sites (Casacuberta and Santiago 2003; Witte et al. 2001; Yang et al. 2007; Lu et al. 2012; Antonius-Klemola et al. 2006; Havecker et al. 2004; Benjak et al. 2009). TRIM elements are actively involved in rearrangements of the highly duplicated *Brassica* genomes. It has been suggested that TRIM insertion into genes has mediated the gain of different functions, expression and evolution of triplicated genes (neofunctionalization) in *Brassica* genomes (Yang et al. 2007). Our previous comparative analysis of *Cassandra* (TB-5)family members in *A. thaliana, B. rapa* and *B. oleracea* suggested that some *Cassandra* elements have been commonly retained during the last 20 million years in the three species and that some elements have uniquely evolved in specific *Brassica* species (Sampath and Yang 2014a, b) (Fig. 6.2a).

Many studies have demonstrated that SINEs play a significant role in plant genetic variation and genomic evolution, including changing gene function and/or expression by providing regulatory elements such as alternative splicing sites and poly adenylation signals for functional RNA genes (Kramerov and Vassetzky 2011a, b; Ben-David et al. 2013). SINEs also cause insertional mutations and alter the methylation pattern in the genome (Batzer and Deininger 2002).

MITEs also play an important role in gene regulation and rearrangement and expression (Sampath et al. 2013, 2014). Transposition of MITEs into genes have been found to modify gene structure and function by deletion, point mutation, and affecting the transcriptional activity (Wessler et al. 1995; Mo et al. 2012; Sarilar et al. 2011; Shirasawa et al. 2012). Specifically, MITE transposition into introns in triplicated *B. rapa* genes appears to underlie their differential expression patterns (Sampath et al. 2013) (Fig. 6.2b). Although most MITEs are associated with genic regions, they are generally not found in exons. An exception is a *tourist* family of MITEs from *B. rapa*, *BraTo-9*, which is preferentially present in the exons of triplicated *B. rapa* genes. *BraTo-9* has provided new exons for functional genes of *B. rapa* (Sampath et al. 2014). When *BraTo-9* insertion occurred in triplicated or duplicated genes of *B. rapa*, the element was always found in

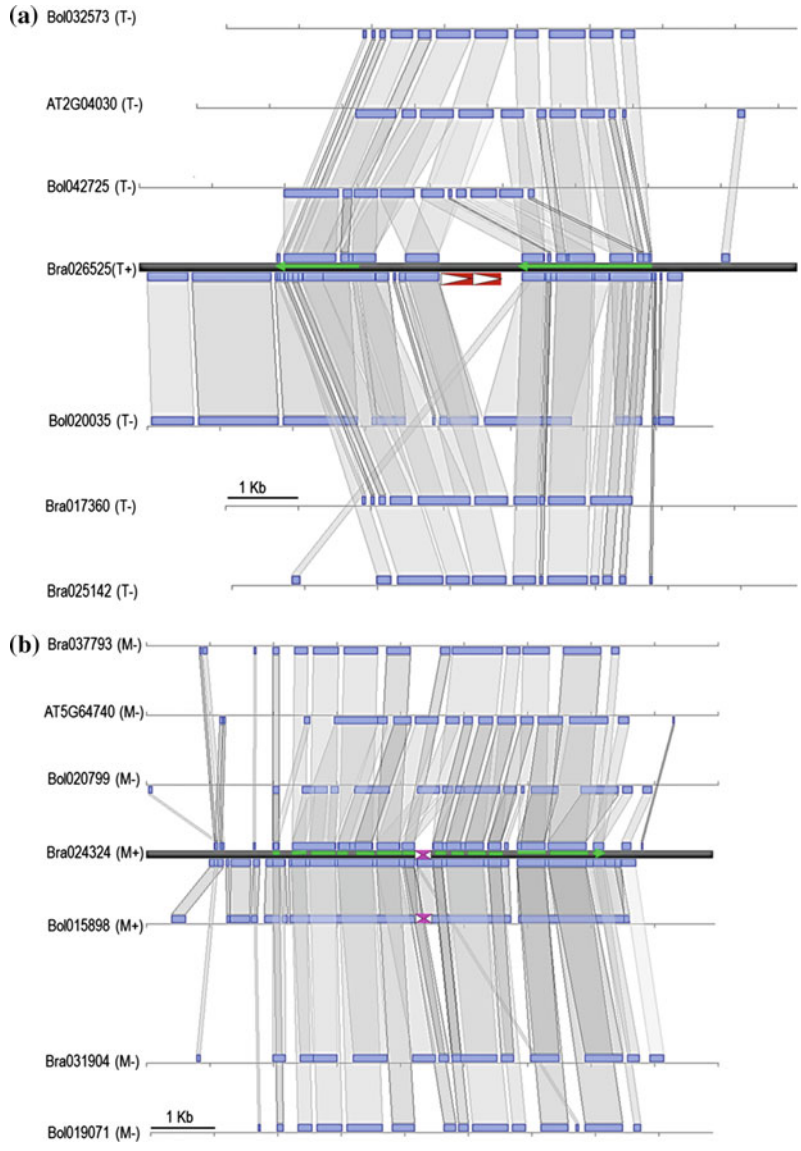

Fig. 6.2 Microsynteny comparison of *B. rapa* genomic regions containing TRIM (*TB-5*) and MITE (*BraSto-2*) elements with their noninserted paralogs (NIPs) and noninserted orthologs (NIOs) in *A. thaliana* and *B. oleracea*. **a** Unique *TB-5* (Br) element insertion in *B. rapa*. Microsynteny comparison of the genomic region of *TB-5* (Br) (Bra026525) with its NIPs (Bra025142, Bra017360) and NIOs from *B. oleracea* (Bol020035, Bol032573, Bol042725) and *A. thaliana* (AT2G04030). **b** Microsynteny between the genomic region showing shared insertion of *BraSto-2* in genes of *B. rapa* (Bra024324) and *B. oleracea* (Bol015898) compared with those of its NIPs of *B. rapa* (Bra037793, Bra031904) and *B. oleracea* (Bol020799, Bol019071) and its NIO from *A. thaliana* (AT5G64740). Exons and gene direction are indicated with green arrows. mTEelement insertions are shown as *red* and *pink bars*. + and − indicate genes with TRIM (T) or MITE (M) insertion and non-insertion, respectively. The *gray bars* connecting boxes on genome sequences indicate syntenic blocks present in both sequences. The map was generated based on nucleotide sequence similarity determined by BLASTn search

only one of the duplicated or triplicated genes, suggesting that the *BraTo-9* members were actively amplified in *B. rapa* after divergence with *B. oleracea* 4.6 MYA. MITE excision has caused gene knockout or silencing, and up- or down-regulation of gene expression by gene

rearrangement, *trans* duplication, and footprint mutation (Benjak et al. 2009; Shirasawa et al. 2012; Naito et al. 2009). In addition, MITEs are sources of small interfering RNA and can control genes in their vicinity (Piriyapongsa and Jordan 2007; Piriyapongsa et al. 2007; Kuang et al. 2009). *Trans* duplication in MITEs (i.e., MITEs with host gene sequence captured during excision) increases the likelihood of generating siRNA, which can influence gene regulation (Benjak et al. 2009; van Leeuwen et al. 2003). For instance, a MITE-based siRNA represses the expression of nearby genes by acting as a functional regulator triggering DNA methylation, and thereby affects agronomic traits such as leaf angle, plant height and inflorescence morphology (Wei et al. 2014). MITEs also have the ability to escape from silencing more efficiently than other TEs (Benjak et al. 2009; Parisod et al. 2010).

6.5 Copy Number Variation of mTE Members Between Accessions and Species

We estimated the copy numbers for each mTE family for each mTE in *B. rapa* compared to those in *B. oleracea* (Table 6.2) based on mapping of 1× coverage whole-genome sequence (WGS) reads using the criteria of 80 % sequence similarity with 80 % coverage against representative members of 39 mTE families. This analysis showed that *B. oleracea* has more mTE copies than does *B. rapa*, with 3-fold differences observed. In addition, copy numbers of some mTEs vary up to 2-fold among accessions of *B. rapa*. Together, these data suggest that mTE members were greatly amplified after the *B. rapa* and *B. oleracea* diversification about 4.6 MYA (Table 6.2; Fig. 6.3) (Mun et al. 2009).

The copy numbers of the five TRIM families (TB-1-5) ranged between 375 and 541 among *B. rapa* accessions, suggesting that TRIM elements have been active recently in different accessions (Fig. 6.4a; Table 6.2). Differential amplification of TRIM elements within or between species will be important target for identification of functional genes associated with mTE insertions and also for molecular breeding purposes (Fig. 6.5). More analysis related to gene function and association of TRIM elements could also promote the identification of agriculturally important genes.

Our analysis based on 14 SINE families shows that SINEs have been differentially amplified, with between 207 and 541 copies in *B. rapa* accessions but a larger range (207–1341) between *B. rapa* and *B. oleracea* (Fig. 6.4b; Table 6.2). Most of the SINE families exhibit similar amplification between both species, but some families (SB-Wicker et al. 2007; Liu et al. 2014; Bire and Rouleux-Bonnin 2012; Feschotte 2008; Alzohairy et al. 2013; Hollister and Gaut 2009; Lisch 2009) are more abundant in *B. oleracea* than in *B. rapa* (Table 6.2). This suggests that SINE elements also have proliferated with recent activation in both genomes, but more so in *B. oleracea* than in *B. rapa* (Fig. 6.5). These

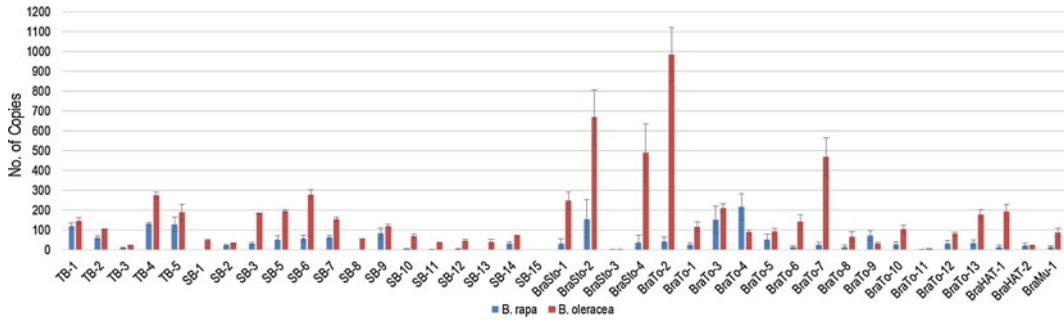

Fig. 6.3 Copy numbers of 39 mTE families in *B. rapa* and *B. oleracea* genomes based on 1× WGS data. The number of mTE members (average) and standard deviation were calculated from 11 *B. rapa* and 2 *B. oleracea* accessions

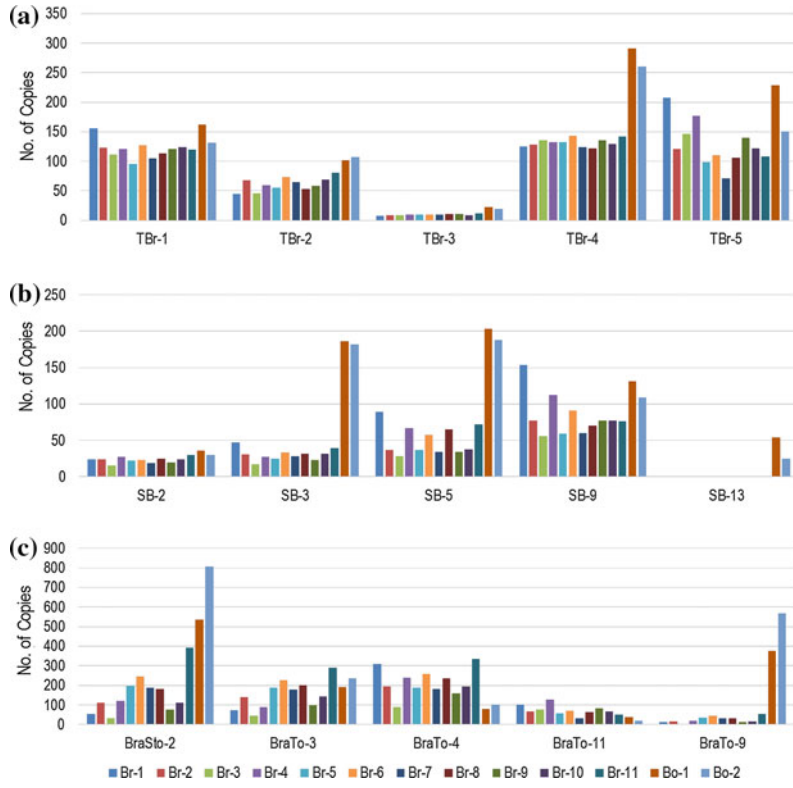

Fig. 6.4 Distribution of mTE family members in *B. rapa* and *B. oleracea* genomes based on 1× WGS data. The *graph* shows the total members from five families of TRIM (**a**), SINE (**b**) and MITE (**c**) elements based on 10 *B. rapa* accessions (Br-1 ~ 10) and two *B. oleracea* accessions (Bo 1, 2). The total copy numbers of mTE families were calculated based on 1× WGS read mapping. The accession names are listed in Table 6.3

diversely amplified members can be important resources for molecular and evolutionary studies.

MITEs also show significant divergence in copy number between *B. rapa* accessions (range 337–1968) and between *B. rapa* and *B. oleracea* (range 337–5076) (Fig. 6.4c; Table 6.2). This indicates that, like the other mTEs, MITEs remained highly active in both *Brassica* genomes. Thus, differential accumulation of MITEs within or between species should also be an important target for molecular breeding purposes and evolutionary studies (Fig. 6.5).

6.6 Utility of mTEs as Molecular Markers

DNA markers are used for a wide range of genomic applications such as construction of genetic linkage maps, genome-wide association studies and evolutionary studies (Casa et al. 2000; Kwon et al. 2007; Purugganan and Wessler 1995; Yaakov et al. 2012). TEs have been used to develop molecular markers such as those for inter-retrotransposon amplified polymorphism (IRAP), REtrotransposn-microsatellite amplified polymorphism (REMAP), sequence-specific amplification polymorphism (S-SAP), retrotransposon-based insertion polymorphism (RBIP), inter-MITE polymorphism (IMP) and transposon display (TD) (Agarwal et al. 2008). TE-based markers have been successfully utilized for various genomics purposes such as analysis of genetic diversity, inspection of clonal variation and breeding. TE markers are also useful to identify unambiguous gene flow between closely related species (Bire and Rouleux-Bonnin 2012; Carrier et al. 2012; Deragon and Zhang 2006). The principle characteristics of mTEs, namely their abundance, small size, stability, and distribution in genic regions, are advantageous for DNA marker development in both plants and animals. Thus, so-called mTE

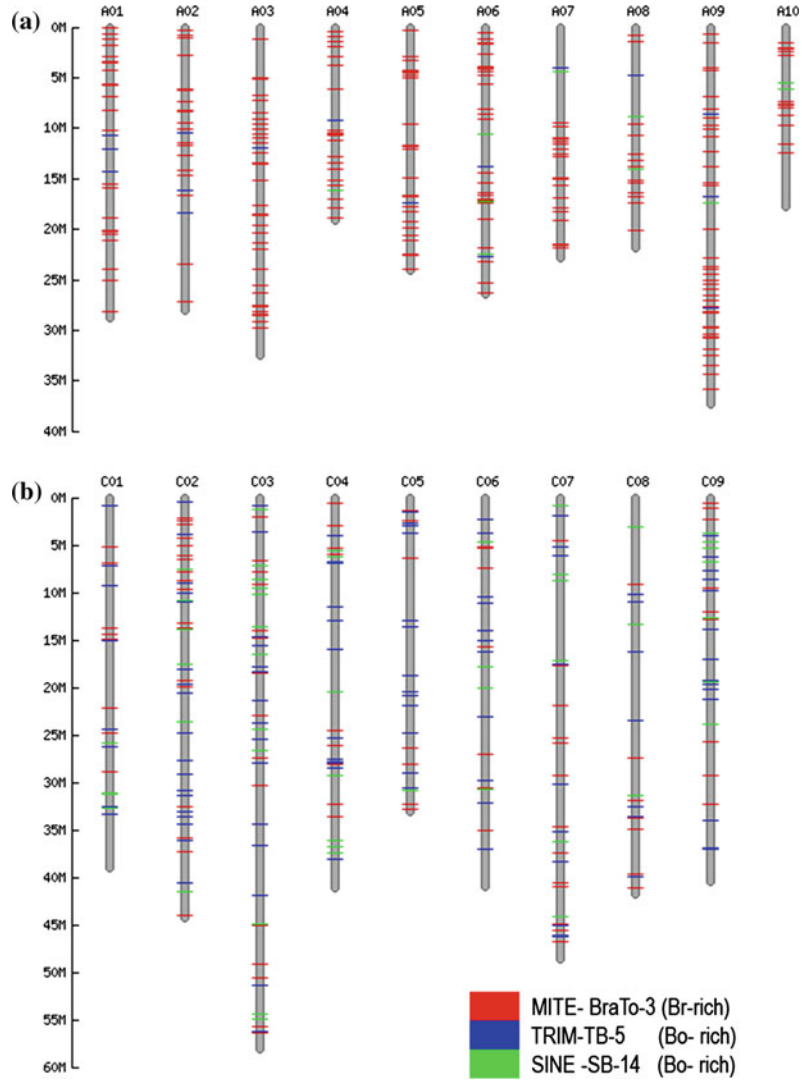

Fig. 6.5 Differential distribution of mTE family members in *B. rapa* and *B. oleracea*. mTE families with intact members were used for in silico map construction on the 256-Mb *B. rapa* (**a**) and the 385-Mb *B. oleracea* (**b**) pseudo-chromosome sequences based on physical positions. The physical position information for the mTE families of *B. rapa* and *B. oleracea* can be found in BrassicTED (24)

markers have been developed from mTEs such as TRIMs (Witte et al. 2001; Kwon et al. 2007), SINEs (Deragon and Zhang 2006; Shedlock and Okada 2000; Tatout et al. 1999) and MITEs (Shirasawa et al. 2012; Casa et al. 2000).

6.6.1 Insertion *Polymorphism of mTEs*

The presence (inserted site) or absence (empty site) of an mTE at a particular locus can be different among accessions, and this insertion polymorphism (IP) can be surveyed (Kwon et al. 2007; Yaakov et al. 2012) by PCR analysis using primers designed from the mTE flanking region (Fig. 6.6a). The mTE markers have an advantage over other types of markers because the stability and high copy numbers of mTEs allows development of abundant markers (Monden et al. 2009). IP markers represent co-dominant alleles at a single locus and can be used for applications such as identification of genome duplication or allopolyploidization events, genetic diversity analysis among related accessions or species, and development of markers and mapping using

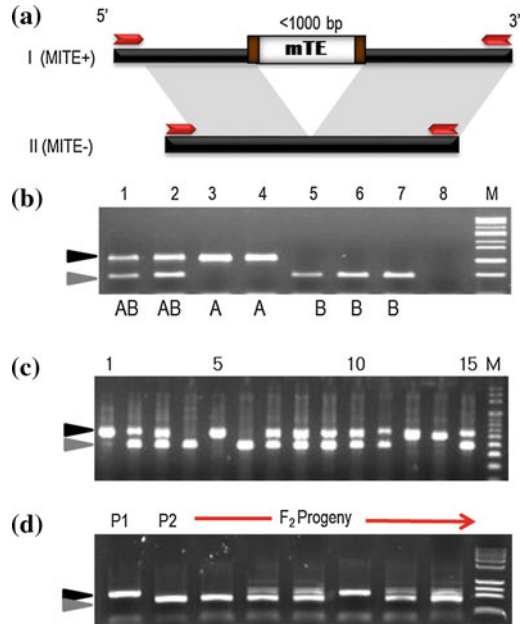

Fig. 6.6 Utility of mTEs as molecular markers**. a** MITE insertion polymorphism analysis using flanking primers. Comparison of DNA fragments showing the presence or absence of MITE insertion. MITE-flanking primer positions are indicated as *red arrowheads*. **b** Polymorphism profile by MIP analysis of 7 *Brassica* accessions based on *BraMi-1*, a *Brassica* MITE. *AB* insertion and non-insertion (Heterozygous insertion); *A* Insertion (Homozygous insertion); *B* non-insertion (Homozygous non-insertion). The list of accessions used and their ploidy are given in Sampath et al. (33). **c** Diversity analysis using different *B. oleracea* commercial cultivars. **d** Genotyping analysis of 94 *B. oleracea* F_2 plants from a cross between parental lines C1234 (P1) and C1184 (P2)

segregating populations between parental lines (Fig. 6.6b–d) (Sampath et al. 2013).

IP markers can be produced from sequences harboring TRIM, SINE, or MITE elements near the genes of interest. TRIM insertion polymorphism (TIP) markers were successfully developed to analyze genetic diversity and insertion time through divergent appearance of various TRIM elements in different *Brassica* accessions (Yang et al. 2007). SINE insertion polymorphism (SIP) markers have been used for construction of genetic maps, candidate-gene association studies and analysis of evolution within the *Glycine* genus. More than 52 % (77/146) of SIP markers developed from members of the GmAu1 mTE family displayed polymorphism (Shu et al. 2011). In addition, MITE insertion polymorphism (MIP) markers have been extensively studied in rice using a *Tourist* family MITE, *mPing*, the first active MITE identified in eukaryotes (Monden et al. 2009). MIP markers based on three MITEs (*Hbr, zmv1, Ins2*) were successfully used to study genetic diversity and identify a new candidate gene for flowering time variation in maize (Casa et al. 2000). MIP markers also have been used for high-resolution genetic diversity analysis and to elucidate the evolutionary history of *Triticum* (Yaakov et al. 2012). A MIP survey of three different *Brassica* accessions revealed high levels of inter- and intra-species polymorphism, at 52 % (150 markers) and 23 % (66 markers), respectively (Sampath et al. 2014). Transposition of MITEs and evolutionary dynamics were also evaluated in *Brassica* species using a MIP approach (Sampath et al. 2013). Thus, mTEs can be valuable targets from which to produce high numbers of successful markers quickly. It is important to note that high IP ratios are dependent on recent activation and high copy numbers for the target mTEs. Our analysis shows that compared with TRIMs and SINEs, MITEs are high copy and differentially amplified inter- and intra-species, suggesting that MITEs are particularly good candidates as targets for plant genome analysis.

6.6.2 Transposon *Display (TD) for mTEs*

TD is a modification of the AFLP method to target TEs and amplify most of the insertion sites of TEs. TD is be an efficient approach for rapid marker development because multiple insertion sites can be simultaneously amplified using conserved sequences of target mTEs that are distributed throughout the genome. TD was first developed and used for the maize *heartbreaker* MITE family (Casa et al. 2004). TD can be performed with primers targeting conserved regions of mTEs, such terminal inverted repeats (TIRs) for MITEs. TD-based markers have been effectively utilized for examining genetic diversity, phylogenetic analysis, genetic mapping,

identification of activation time of TEs based on divergence time and evolutionary studies (Kwon et al. 2007; Monden et al. 2009; Naito et al. 2006). mTE-based display, termed mTE-TD, has been applied for genome-wide detection of insertion sites that are polymorphic between or within species such as rice, maize, *Brassica*, *Vitis vinifera* (grapevine) and mosquito (Naito et al. 2006, 2009; Kwon et al. 2007; Zhang et al. 2000). The mTE-TD approach has advantages over AFLP because mTEs are more widely distributed in genome, especially in euchromatin regions. In addition, mTEs are closely associated with genic regions, which may help to develop markers related to agronomically important traits. Reports have also suggested that mTE-TD identifies a higher proportion of polymorphisms than does AFLP. TRIM-TD using conserved regions of *TRIM-Br1&2* revealed various insertion sites and were used for genetic mapping and high-resolution genetic diversity analysis of various *Brassica* relatives (Kwon et al. 2007).

Next-generation sequencing (NGS) technology produces numerous short DNA reads at relatively low cost and in a short period of time. NGS has a wide range of applications and has revolutionized the use of genomic data for crop improvement (Wei et al. 2013). The combination of TD with NGS technology allows the use of different high copy mTE families to detect insertion polymorphism among accessions. This approach will be a powerful tool for molecular breeding and evolutionary analysis.

6.7 Conclusion

Although mTEs cannot transpose by themselves due to their lack of protein-coding genes, mTEs have played important roles in plant genome evolution. Understanding the characteristics and member distribution of mTEs will promote their effective utilization to analyze genome evolution, dynamics, and plasticity as well as to identify the relevant genetic components of germplasm with agronomically important traits in the *Brassica* genome. The mTE-based markers are valuable resources for high-density genetic mapping, diversity analysis and evolution studies. Furthermore, insertion polymorphism surveys and NGS combined with TD are potential tools for marker systems aimed at high throughput marker development with minimum time and cost.

Acknowledgments This research was carried out with the support by Golden Seed Project (Center for Horticultural Seed Development, No. 213003-04-3-SB430), Ministry of Agriculture, Food and Rural Affairs (MAFRA), Ministry of Oceans and Fisheries (MOF), Rural Development Administration (RDA) and Korea Forest Service (KFS), Republic of Korea.

References

Agarwal M, Shrivastava N, Padh H (2008) Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep 27(4):617–631

Alzohairy AM, Gyulai G, Jansen RK, Bahieldin A (2013) Transposable elements domesticated and neofunctionalized by eukaryotic genomes. Plasmid 69(1):1–15

Antonius-Klemola K, Kalendar R, Schulman AH (2006) TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. TAG Theor Appl Genet Theoretische und angewandte Genetik 112(6):999–1008

Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3(5):370–379

Ben-David S, Yaakov B, Kashkush K (2013) Genome-wide analysis of short interspersed nuclear elements SINES revealed high sequence conservation, gene association and retrotranspositional activity in wheat. Plant J Cell Mol Biol 76(2):201–210

Benjak A, Boue S, Forneck A, Casacuberta JM (2009) Recent amplification and impact of MITEs on the genome of grapevine (*Vitis vinifera* L.). Genome Biol Evol 1:75–84

Bennett MD, Smith J (1976) Nuclear DNA amounts in angiosperms. Philos Trans R Soc Lond B Biol Sci 274 (933):227–274

Bire S, Rouleux-Bonnin F (2012) Transposable elements as tools for reshaping the genome: it is a huge world after all! Methods Mol Biol 859:1–28

Bureau TE, Wessler SR (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4(10):1283–1294

Bureau TE, Wessler SR (1994) Stowaway: a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6(6):907–916

Carrier G, Le Cunff L, Dereeper A, Legrand D, Sabot F, Bouchez O et al. (2012) Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L. PloS One**7**(3):e32973

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S et al (2000) The MITE family heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 97(18):10083–10089

Casa AM, Nagel A, Wessler SR (2004) MITE display. Methods Mol Biol 260:175–188

Casacuberta JM (2013) MITEs, miniature elements with a major role in plant genome evolution. Plant Trans Elem Impact Genome Struct Funct 24:113

Casacuberta JM, Santiago N (2003) Plant LTR-retrotransposons and MITEs: control of transposition and impact on the evolution of plant genes and genomes. Gene 311:1–11

Chen Y, Zhou F, Li G, Xu Y (2009) MUST: a system for identification of miniature inverted-repeat transposable elements and applications to Anabaena variabilis and Haloquadratum walsbyi. Gene 436(1–2):1–7

Chen J, Hu Q, Zhang Y, Lu C, Kuang H (2013) P-MITE: a database for plant miniature inverted-repeat transposable elements. Nucleic acids research 42(D1): D1176–D1181

Deragon J-M, Zhang X (2006) Short interspersed elements (SINEs) in plants: origin, classification, and use as phylogenetic markers. Syst Biol 55(6):949–956

Fattash I, Rooke R, Wong A, Hui C, Luu T, Bhardwaj P et al (2013) Miniature inverted-repeat transposable elements: discovery, distribution, and activity1. Genome Natl Res Counc Can Genome Conseil Natl de Rech Can 56(9):475–486

Feschotte C (2008) Transposable elements and the evolution of regulatory networks. Nat Rev Genet 9 (5):397–405

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Ann Rev Genet 41:331

Feschotte C, Wessler SR (2002) Mariner-like transposases are widespread and diverse in flowering plants. Proc Natl Acad Sci USA 99(1):280–285

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3(5):329–341

Han Y, Wessler SR (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res 38(22):e199

Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. Genome Biol 5(6):225

Hollister JD, Gaut BS (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. Genome Res 19(8):1419–1428

Janicki M, Rooke R, Yang G (2011) Bioinformatics and genomic analysis of transposable elements in eukaryotic genomes. Chromosome Res Int J Mol Supramol Evolutionary Aspects Chromosome Biol 19(6):787–808

Jurka J (2000) Repbase update: a database and an electronic journal of repetitive elements. Trends Genetics 16(9): 418–420

Kalendar R, Tanskanen J, Chang W, Antonius K, Sela H, Peleg O et al (2008) Cassandra retrotransposons carry independently transcribed 5S RNA. Proc Natl Acad Sci USA 105(15):5833–5838

Kramerov DA, Vassetzky NS (2005) Short retroposons in eukaryotic genomes. Int Rev Cytol 247:165–221

Kramerov DA, Vassetzky NS (2011) Origin and evolution of SINEs in eukaryotic genomes. Heredity 107 (6):487–495

Kramerov DA, Vassetzky NS (2011b) SINEs. Wiley interdisciplinary reviews. RNA 2(6):772–786

Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, et al.(2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the Solanaceae: new functional implications for MITEs. Genome Res 19 (1):42–56

Kwon SJ, Kim DH, Lim MH, Long Y, Meng JL, Lim KB et al (2007) Terminal repeat retrotransposon in miniature (TRIM) as DNA markers in *Brassica* relatives. Mol Genet Genomics MGG 278(4):361–370

Lerat E (2009) Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. Heredity 104(6):520–533

Lisch D (2009) Epigenetic regulation of transposable elements in plants. Ann Rev Plant Biol 60:43–66

Lisch D (2012) Regulation of transposable elements in maize. Curr Opin Plant Biol 15(5):511–516

Liu S, Liu Y, Yang X, Tong C, Edwards D, Parkin IA et al (2014) The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun 5:3930

Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H (2012) Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. Mol Biol Evol 29 (3):1005–1017

Mo Y-J, Kim K-Y, Shin W-C, Lee G-M, Ko J-C, Nam J-K et al.(2012) Characterization of Imcrop, a Mutator-like MITE family in the rice genome. Genes Genomics 34(2):189–198

Monden Y, Naito K, Okumoto Y, Saito H, Oki N, Tsukiyama T et al.(2009) High potential of a transposon mPing as a marker system in japonica x japonica cross in rice. DNA Res Int J Rapid Publ Rep Genes Genomes 16(2):131–140

Mun JH, Kwon SJ, Yang TJ, Seol YJ, Jin M, Kim JA et al.(2009) Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. Genome Biol 10 (10):R111

Murukarthick J, Sampath P, Lee SC, Choi BS, Senthil N, Liu S et al.(2014) BrassicaTED—a public database for utilization of miniature transposable elements in *Brassica* species. BMC Res Notes **7**(1):379

Naito K, Cho E, Yang G, Campbell MA, Yano K, Okumoto Y et al (2006) Dramatic amplification of a rice transposable element during recent domestication. Proc Natl Acad Sci USA 103(47):17620–17625

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO et al. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461 (7267):1130–1134

Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ et al (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector. Sci Signal 316(5832):1718

Okada N, Hamada M, Ogiwara I, Ohshima K (1997) SINEs and LINEs share common 3′ sequences: a review. Gene 205(1):229–243

Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T (2008) A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. japonica. Genes Genet Sys 83 (4):321–329

Parisod C, Alix K, Just J, Petit M, Sarilar V, Mhiri C et al (2010) Impact of transposable elements on the organization and function of allopolyploid genomes. New Phytol 186(1):37–45

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H et al. (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457(7229):551–556

Piriyapongsa J, Jordan IK (2007) A family of human microRNA genes from miniature inverted-repeat transposable elements. PloS One 2(2):e203

Piriyapongsa J, Marino-Ramirez L, Jordan IK (2007) Origin and evolution of human microRNAs from transposable elements. Genetics 176(2):1323–1337

Purugganan MD, Wessler SR (1995) Transposon signatures: species-specific molecular markers that utilize a class of multiple-copy nuclear DNA. Mol Ecol 4 (2):265–269

Sampath P, Yang T-J (2014) Comparative analysis of cassandra TRIMs in three *Brassicaceae* genomes. Plant Genet Resour Charact Utilization 12(S1):S146–S150

Sampath P, Yang TJ (2014b) Miniature inverted-repeat transposable elements (MITEs) as valuable genomic resources for evolution and breeding of Brassica crops. Plant Breed Biotechnol 2:322–333

Sampath P, Lee S-C, Lee J, Izzah NK, Choi B-S, Jin M et al. (2013) Characterization of a new high copy stowaway family MITE, BRAMI-1 in *Brassica* genome. BMC Plant Biol 13(1):56

Sampath P, Murukarthick J, Izzah NK, Lee J, Choi HI, Shirasawa K et al. (2014) Genome-wide comparative analysis of 20 miniature inverted-repeat transposable element families in *Brassica rapa* and *B. oleracea*. PloS One 9(4):e94499

Sarilar V, Marmagne A, Brabant P, Joets J, Alix K. BraSto (2011) A stowaway MITE from *Brassica*: recently active copies preferentially accumulate in the gene space. Plant Mol Biol 77(1–2):59–75

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W et al. (2010) Genome sequence of the palaeopolyploid soybean. Nature 463(7278):178–183

Shedlock AM, Okada N (2000) SINE insertions: powerful tools for molecular systematics. BioEssays News Rev Mol Cell Dev Biol 22(2):148–160

Shirasawa K, Hirakawa H, Tabata S, Hasegawa M, Kiyoshima H, Suzuki S et al (2012) Characterization of active miniature inverted-repeat transposable elements in the peanut genome. Theor Appl Genet 124 (8):1429–1438

Shu Y, Li Y, Bai X, Cai H, Ji W, Ji Z et al. (2011) Identification and characterization of a new member of the SINE Au retroposon family (GmAu1) in the soybean, *Glycine max* (L.) Merr., genome and its potential application. Plant Cell Rep 30(12):2207–2213

Smit A, Hubley R (1996) Repeat modeler open- 3.0. Available at http://www.repeatmasker.org/

Tatout C, Warwick S, Lenoir A, Deragon J-M (1999) SINE insertions as clade markers for wild crucifer species. Mol Biol Evol 16(11):1614

Tu Z (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, Anopheles gambiae. Proc Natl Acad Sci 98 (4):1699–1704

van Leeuwen H, Monfort A, Zhang H-B, Puigdomènech P (2003) Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between *Cucumis melo* and *Arabidopsis thaliana*. Plant Mol Biol 51(5):703–718

Vassetzky NS, Kramerov DA (2013) SINEBase: a database and tool for SINE analysis. Nucleic Acids Res 41(Database issue):D83–D89

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG et al. (2001) The sequence of the human genome. Science 291(5507):1304–1351

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S et al. (2011) The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet 43(10):1035–1039

Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G (2004) Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res 14 (10a):1861–1869

Wei L, Xiao M, Hayward A, Fu D (2013) Applications and challenges of next-generation sequencing in *Brassica* species. Planta 238:1005–1024

Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M et al (2014) Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. Proc Natl Acad Sci USA 111(10):3877–3882

Wessler SR (2006) Transposable elements and the evolution of eukaryotic genomes. Proc Natl Acad Sci USA 103(47):17600–17601

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players

in the evolution of plant genomes. Curr Opin Genet Dev 5(6):814–821

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B et al (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973–982

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M et al (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59 (5):712–722

Witte C-P, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci 98(24):13778–13783

Yaakov B, Kashkush K (2012) Mobilization of Stowaway-like MITEs in newly formed allohexaploid wheat species. Plant Mol Biol 80(4–5):419–427

Yaakov B, Ceylan E, Domb K, Kashkush K (2012) Marker utility of miniature inverted-repeat transposable elements for wheat biodiversity and evolution. TAG Theor Appl Genet Theoretische und angewandte Genetik 124(7):1365–1373

Yang G (2013) MITE Digger, an efficient and accurate algorithm for genome wide discovery of miniature inverted repeat transposable elements. BMC Bioinfor 14:186

Yang TJ, Kwon SJ, Choi BS, Kim JS, Jin M, Lim KB et al. (2007) Characterization of terminal-repeat retrotransposon in miniature (TRIM) in *Brassica* relatives. TAG Theor Appl Genet Theoretische und angewandte Genetik 114(4):627–636

Zhang Q, Arbuckle J, Wessler SR (2000) Recent, extensive, and preferential insertion of members of the miniature inverted-repeat transposable element family Heartbreaker into genic regions of maize. Proc Natl Acad Sci USA 97(3):1160–1165