Mol Genet Genomics (2015) 290:2297–2312
DOI 10.1007/s00438-015-1076-9

ORIGINAL PAPER

Evolutionary genomics of miniature inverted-repeat transposable elements (MITEs) in *Brassica*

Faisal Nouroz[1,3] · Shumaila Noreen[2] · J. S. Heslop-Harrison[1]

Received: 23 December 2014 / Accepted: 29 May 2015 / Published online: 1 July 2015

Abstract Miniature inverted-repeat transposable elements (MITEs) are truncated derivatives of autonomous DNA transposons, and are dispersed abundantly in most eukaryotic genomes. We aimed to characterize various MITEs families in *Brassica* in terms of their presence, sequence characteristics and evolutionary activity. Dot plot analyses involving comparison of homoeologous bacterial artificial chromosome (BAC) sequences allowed identification of 15 novel families of mobile MITEs. Of which, 5 were Stowaway-like with TA Target Site Duplications (TSDs), 4 Tourist-like with TAA/TTA TSDs, 5 Mutator-like with 9–10 bp TSDs and 1 novel MITE (*BoXMITE1*) flanked by 3 bp TSDs. Our data suggested that there are about 30,000 MITE-related sequences in *Brassica rapa* and *B. oleracea* genomes. In situ hybridization showed one abundant family was dispersed in the A-genome, while another was located near 45S rDNA sites. PCR analysis using primers flanking sequences of MITE elements detected MITE insertion polymorphisms between and within the three *Brassica* (AA, BB, CC) genomes, with many insertions being specific to single genomes and others showing evidence of more recent evolutionary insertions. Our BAC sequence comparison strategy enables identification of evolutionarily active MITEs with no prior knowledge of MITE sequences. The details of MITE families reported in *Brassica* enable their identification, characterization and annotation. Insertion polymorphisms of MITEs and their transposition activity indicated important mechanism of genome evolution and diversification. MITE families derived from known Mariner, Harbinger and Mutator DNA transposons were discovered, as well as some novel structures. The identification of *Brassica* MITEs will have broad applications in *Brassica* genomics, breeding, hybridization and phylogeny through their use as DNA markers.

Keywords Biodiversity · *Brassica* · Genome evolution · Genomics · MITEs · Transposable elements

Introduction

Miniature inverted-repeat transposable elements (MITEs) represent a heterogeneous group of short non-autonomous mobile DNA elements, considered to be deletion derivates of autonomous DNA transposons. The mechanism of amplification and insertion into genomic DNA causes these elements to be flanked by Target Site Duplications (TSDs), and the elements themselves have Terminal Inverted Repeats (TIRs) of varying lengths that are typically AT-rich. MITEs are short (typically 70–500 bp long) and have been found in high copy numbers in plant (Wessler et al. 1995), fungal (Bergemann et al. 2008) and animal (Kuhn and Heslop-Harrison 2011) genomes. Initially two MITE groups were recognized based on their structural features: Stowaway-like MITEs with TA TSDs and Tourist-like MITEs with TAA TSDs (Jiang

Communicated by M.-A. Grandbastien.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-1076-9) contains supplementary material, which is available to authorized users.

✉ Faisal Nouroz
faisalnouroz@gmail.com

[1] Molecular Cytogenetics Laboratory, Department of Biology, University of Leicester, Leicester LE1 7RH, UK

[2] Behavioural Genetics Laboratory, Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

[3] Present Address: Department of Botany, Hazara University Mansehra, Mansehra 21300, Pakistan

et al. 2004b; Zhang et al. 2004). A 128 bp MITE insertion in a mutant maize waxy gene was the first Tourist element discovered (Bureau and Wessler 1992), while a 257 bp element from sorghum was the first element of the Stowaway MITE group (Bureau and Wessler 1994). Later on, MITEs derived by sequence deletion from other DNA transposon families including hAT, CACTA, Mutator, and PiggyBac have been identified in numerous species (Benjak et al. 2009; Kuang et al. 2009; Menzel et al. 2014), making non-autonomous short elements that lack any protein coding domains but retain some sequence characteristics (TSDs and TIRs) of their progenitors enabling classification (Feschotte and Mouches 2000; Yang and Hall 2003; Jiang et al. 2004a, b). As derivatives of active DNA transposons lacking the transposase protein are necessary for their transposition and integration in new sites, mobility of MITEs is enabled by genes in their autonomous partners (Wicker et al. 2007). MITEs are mostly located in euchromatic (gene rich) regions and have played a role in gene regulation, in few cases downregulating the expression of nearby genes (Lu et al. 2012). The role of MITEs in modulation of gene expression and diversification of important crops such as barley (Lyons et al. 2008; Sun et al. 2009) and rice (Lu et al. 2012), as well as in animals such as *Xenopus* (Hikosaka et al. 2011), is now established. Sun et al. (2009) investigated insertion of MITEs into low-copy-number sequences or genic regions in rice, finding that a nested structure (along with other deletion/insertions) has modified the *RPB2* (nuclear RNA polymerase II) gene, and is useful for phylogenetic and phylogeographic analyses in the *Hordeum/Elymus* group of species in the Triticeae.

The present study aimed to identify MITE-related sequences derived from various DNA superfamilies in *Brassica* genomes using dot plot comparisons of homoeologous BAC sequences and BLASTN searches. MITE-like sequences were identified and classified by their TSDs, TIRs, overall length and lack of open reading frames to gain an understanding of the diversity of small non-autonomous mobile elements in *Brassica*. We aimed to analyse MITEs to identify and describe characteristics of known and any novel *Brassica* MITE families, and then aimed to examine abundance, mobility, amplification and evolutionary history of the elements in various *Brassica* species and multiple diverse accessions to identify the contribution of individual mobile MITEs to the diversity of *Brassica*.

Materials and methods

Dot plot and BLASTN approaches for *Brassica* MITE identification

Pairs of *Brassica rapa* (AA) and *B. oleracea* (CC) BAC sequences (Supplementary Fig.) with long stretches of homoeology were identified by dot plots [(using JDotter (Sonnhammer and Durbin 1995) and Dotlet (Junier and Pagni 2000)] programs. Deletion–insertion pairs where one BAC had a sequence fragment that was absent from the other homologue were identified. The junction points in the sequence with the insertion were manually examined for evidence of TSDs and TIRs. Dot plots were also used to identify MITEs with long TIRs (MITEs derived from Mutator-like elements) within a *B. rapa* or *B. oleracea* BAC sequence plotted against itself. Homology lines perpendicular and close to the principal diagonal line of homology showed the TIRs, which were further investigated for MITE characteristics. Elements without MITE characteristics, including hAT/CACTA, full transposons and retrotransposons were also detected in the screen but are not analysed here. BLASTN searches (Altschul et al. 1997, 2009) were used to collect MITE homologues/copies from NCBI *Brassica* nucleotide (nr/nt) collection database (http://www.ncbi.nlm.nih.gov) using dot plot identified (reference) MITEs sequences as query.

MITEs copy number estimation

We analysed MITEs from 62 Mbp of *Brassica* genomic DNA BAC sequences available in the GenBank nucleotide database (nr/nt collection) at NCBI before March, 2012. The numbers of hits against the reference queries with >70 % query coverage and identity were extrapolated after getting output from BLASTN with copy no. = no. in database × genome size/database size as used for the estimation of MITEs in African mosquitoes (Tu 2001). We did not use shotgun-sequenced whole genomic assemblies, as MITEs, particularly when heterozygous, may be omitted.

MITEs characterization and nomenclature

The MITE-like sequences identified in dot plots were BLASTed against Repbase (Jurka et al. 2005) and TIGR Plants Repeat Database (JCVI) (Ouyang and Buell 2004) for homology-based characterization. Names to elements were given systematically: e.g. *BrSTOW1-1*, where '*B*' stands for genus *Brassica*, second letter (*r/o*) represents species name (*rapa/oleracea*), 4 capital letters (*STOW/TOUR*) indicate *Stowaway/Tourist* origin of the MITE, the first number indicates the family and number followed by hyphen represents the respective member of that family as recommended for the nomenclature of transposable elements by Capy (2005). For Mutator-like MITEs such as *BrMuMITE1-1*, '*Mu*' represents Mutator and in MITEs, whose autonomous counterpart was unclear or TSDs/TIRs were ambiguous were described as unknown MITE families such as *BrXMITE1-1*, where '*X*' indicates unknown MITE.

Table 1 *Brassica* accessions used for the study of diversity of MITE elements

No.	Accession	Species	Genomic constitution	Accession name	Crop name
1.	HRIGRU 2488	*B. rapa chinensis*	AA	Pak Choy	Chinese cabbage
2.	HRIGRU 2741	*B. rapa pekinensis*	AA	Chinese Wong Bok	Chinese cabbage
3.	HRIGRU 7574	*B. rapa chinensis*	AA	San Yue Man	Pak choi
4.	HRIGRU 11698	*B. rapa rapa*	AA	Hinona	Turnip
5.	HRIGRU 13174	*B. rapa rapa*	AA	Vertus	Turnip
6.	ND	*B. rapa*	AA	Suttons	Turnips (Snow balls)
7.	HRIGRU011011	*B. nigra*	BB	ND	Wild species
8.	HRIGRU010978	*B. nigra*	BB	ND	Wild species
9.	HRIGRU010919	*B. nigra*	BB	ND	ND
10.	PK-001722	*B. juncea*	AABB	NARC-I	ND
11.	ND	*B. juncea*	AABB	NATCO	ND
12.	PK-001325	*B. juncea*	AABB	NARC-II	ND
13.	HRIGRU 2203	*B. oleracea gemmifera*	CC	De Rosny	Brussels sprout
14.	HRIGRU 5108	*B. oleracea*	CC	Kai Lan	ND
15.	HRIGRU2859	*B. oleracea*	CC	Early Snowball	Cauliflower
16.	HRIGRU 7518	*B. oleracea italic*	CC	Precoce Di Calabria Tipo Esportazione	Broccoli
17.	HRIGRU 3211	*B. oleracea capitata*	CC	Cuor Di Bue Grosso	Cabbage
18.	GK97361	*B. oleracea*	CC	ND	ND
19.	HRIGRU 2487	*B. juncea*	AABB	Kai Choy	Mustard cabbage
20.	HRIGRU 7563	*B. juncea*	AABB	Megarrhiza	Large rooted mustard
21.	HRIGRU 11702	*B. juncea*	AABB	Tsai Sim	Chinese mustard
22.	HRIGRU 11974	*B. juncea*	AABB	W3	Indian oilseed
23.	HRIGRU 12818	*B. juncea*	AABB	Giant Red Mustard	Japanese greens
24.	ND	*B. juncea*	AABB	Varuna	ND
25.	HRIGRU 11967	*B. napus*	AACC	New	Hakuran
26.	HRIGRU 12685	*B. napus oleifera*	AACC	Mar	Oilseed rape
27.	HRIGRU 12800	*B. napus biennis*	AACC	Last and Best	Kale
28.	HRIGRU 13554	*B. napus*	AACC	Fortune	Swede
29.	ND	*B. napus*	AACC	Drakker	ND
30.	ND	*B. napus*	AACC	Tapidor	ND
31.	HRIGRU 2485	*B. carinata*	BBCC	Addis Aceb	Ethiopian mustard
32.	HRIGRU 6232	*B. carinata*	BBCC	Patu	Ethiopian mustard
33.	HRIGRU 6986	*B. carinata*	BBCC	Tamu Tex-sel Greens	Ethiopian mustard
34.	HRIGRU 13160	*B. carinata*	BBCC	Mbeya Green	Ethiopian mustard
35.	R.G.F 32275	*B. carinata*	BBCC	Aworks-67	ND
36.	PK-0085490	*B. carinata*	BBCC	NARC-PK	ND
37.	ND	*B. napus* × *B. nigra*	AABBCC	ND	ND
38.	ND	*B. carinata* × *B. rapa*	AABBCC	ND	ND
39.	ND	*B. napus* × *B. nigra*	AABBCC	ND	ND
40.	ND	*B. napus* × *B. nigra*	AABBCC	ND	ND

ND not determined

PCR amplification of *Brassica* MITEs from diverse accessions

DNA from 40 *Brassica* accessions (Table 1) was used in the present study. Seeds from 32 *Brassica* accessions were kindly provided by Drs Graham Teakle and Guy Barker (Warwick Research Institute (WRI), Warwick, UK; see Walley et al. 2012). Two *B. juncea* and a *B. carinata* accession were collected from the National Agriculture and Research Center (NARC), Islamabad, Pakistan. Seeds for one commercial variety *B. juncea* (NATCO) accession were bought from an Asian supermarket at Leicester. The

DNA from four synthetic allohexaploids ($2n = 6x$) *Brassica* (Ge et al. 2009) was provided by Dr. Xian Hong Ge (University of Wuhan, China). DNA was extracted from young leaves with a standard CTAB method (Doyle and Doyle 1990) and used for PCR amplification. Oligonucleotide primers were designed from the regions flanking MITE insertions using Primer3 (http://frodo.wi.mit.edu/primer3/). PCR amplifications were performed using 50–75 ng *Brassica* genomic DNA in a 15 µl reaction mix containing 2 µl PCR buffer (Kappa, UK), 1.0 mM additional $MgCl_2$, 1 U Taq DNA polymerase (Kappa, UK), 200–250 mM dNTPs and 0.75 µl (10 pmol) of each primer. The thermal cycling conditions were as follows: 3-min denaturation at 94 °C; 35 cycles of 45 s denaturation at 94 °C, 45 s annealing at 52–64 °C (depending on primers) and 1-min extension at 72 °C; a final 3-min extension at 72 °C. PCR products were separated by electrophoresis in 1 % w/v agarose gels with TAE buffer, stained with addition of 1–2 µl ethidium bromide (10 mg/ml) for the detection of DNA bands under UV illumination, and photographed.

Fluorescent in situ hybridization

Seeds were geminated for 2–3 days and root tips were used for the preparation of mitotic chromosomes. The complete MITEs including the flanking regions were amplified by PCR and cleaned after gel electrophoresis (Qiagen). DNA was labelled with digoxigenin-11-dUTP or biotin-11-dUTP by random primer labelling (Invitrogen Bioprime) and used as probes. FISH of *Brassica* chromosomes was performed according to the protocol of Schwarzacher and Heslop-Harrison (2000). The probe mixture contained 50 % (v/v) formamide, 20 % (w/v) dextran sulphate, 2 × SSC, 25–100 ng probe, 20 mg of salmon sperm DNA, 0.3 % SDS (sodium dodecyl sulphate), and 0.12 mM EDTA (ethylenediaminetetraacetic acid). Hybridization and washing were carried out at low stringency (0.1 × SSC at 42 °C). Chromosomes were counterstained with 0.2 mg/ml DAPI (4′, 6-diamidino-2-phenylindole) diluted in McIlvaine's buffer (pH7) and mounted in antifade solution (Citifluor). Examination of slides was carried out with a Zeiss epifluorescence microscope with single band pass filters and equipped with a CCD camera (Optronics, model S97790). The images were overlayed and optimized in Adobe Photoshop CS using only functions affecting the whole of the cropped image equally.

MITE sequences analysis

GC and AT contents of the MITEs were calculated using online available GC-Calculator (http://www.genomicsplace.com/gc_calc.html). Pictograms or logos of the sequence domains were generated with WebLogo (http://weblogo.berkeley.edu/logo.cgi).

Results

Identification and characterization of 15 MITE families in *Brassica*

A total of 14 distinctive MITE families, with mobility in the evolutionary period since separation of *B. rapa* and *B. oleracea* were identified by dot plot sequence comparisons (Table 2a). An additional MITE family was identified in dot plot analysis by the presence of TIRs within a single BAC (Table 2b). Based on the structural features (TSDs and TIRs) of the known DNA transposons, 14 of the *Brassica* MITEs were characterized as being derived from Mariner (Stowaway), PIF/Harbinger (Tourist) and Mutator (MuMITE) elements. The derivation of one MITE family exhibiting 3 bp TSDs was not classified due to the lack of any clear marks or strong homology to any known MITE or autonomous partner; we named this exception *BoXMITE1*. After initial identification of MITEs by dot plot analyses, these MITEs were used as query in BLASTN searches against the *Brassica* nucleotide collection (nr/nt) Genbank database to collect all homologues (Table 2c) in database. A total of 33 Stowaway, 35 Tourist, 27 Mutator and 5 elements of the novel *BoXMITE1* family were chosen for analysis in the 62 Mbp of sequenced BACs. Based on number of MITEs identified in dot plot analysis and BLASTN searches (Table 2a–c), MITEs in whole genome sequences were estimated (Table 3). Structural characteristics including TSDs [TA in Stowaway (Mariner-derived), TAA or TTA in Tourist (PIF/Harbinger-derived), 9 or 10 bp (Mutator-derived), and TTC (*BoXMITE1*)], short lengths, lack of open reading frames, AT richness (ranging to 80 %, although notably the *BoXMITE1* sequence was not AT-rich at 53 %), and high copy numbers confirmed all elements as MITEs (Table 3).

Dot plots (Fig. 1) show the complex range of structures and lack of conservation in length of the TIRs of representatives of the major MITE families. In one Tourist element and two Mutator-derived elements, the TIRs (boxed) span much of the length of the element. Notably, many MITEs included short near duplications within their internal regions, sometimes (*BrTOUR3*, *BoMuMITE3*, *BoXMITE1*) but not in all cases (*BrSTOW1*, *BrTOUR2*, *BoMuMITE4*) representing fragments of the TIRs.

We found element-specific motifs in the TIRs of the five *Brassica* Stowaway and four Tourist MITE families immediately following the TSD insertion site (Fig. 2). For entry into the program, TIRs (Table 2) of each family were aligned to the same length using CLUSTALW with default parameters, and for a very small number of sequences, single bp insertions or extensions were removed (see Table 3 for length ranges): *BrSTOW1* and *BoTOUR4* had short insertions relative to the other family members, while 4 bp

Table 2 MITEs identified from *Brassica* BAC sequences

Element name	BAC Accession	Species	Size	TSDs	Terminal inverted repeats (TIRs)	AT %	Group
a. Identified in dot plots by insertion–gap pairs (present in one species, absent in the other)							
Comparison of *B. rapa* BAC AC189298 against *B. oleracea* BAC EU642504.1							
BoSTOW3-1	EU642504.1	*B. oleracea*	237	TA	AGAGCATCTTTACCG	58	Stowaway
BrTOUR3-1	AC189298.1	*B. rapa*	258	TTA	GGACATCTCCA—(105)	67	Tourist
BoXMITE1-1	EU642504.1	*B. oleracea*	402	TTC	GGCCATGTTCGTTTACGTGTCGCGCGAC-CTACGACCTGCGAC	48	BoXMITE
Comparison of *B. rapa* BAC CU984545 against *B. oleracea* BAC EU579455.1							
BrMuMITE1-1	CU984545.1	*B. rapa*	551	TATCCTATT	122/125	78	Mutator
BrMuMITE2-1	CU984545.1	*B. rapa*	905	CTTTAGAAAC	427/435	81	Mutator
BoMuMITE4-2	EU579455.1	*B. oleracea*	766	TTGGaTtGT	358/351	77	Mutator
Comparison of *B. rapa* BAC AC155344.1 against *B. oleracea* BAC AC240081.1							
BrSTOW1-1	AC155344.1	*B. rapa*	580	TA	TACCTTTCTGTTCCTAAATATAAGATGTTT	76	Stowaway
BoSTOW2-1	AC240081.1	*B. oleracea*	448	TA	GGCGCTAGTCG*	70	Stowaway
BrTOUR1-1	AC155344.1	*B. rapa*	413	TTA	GGGGGTGTTAGTGGGA	73	Tourist
BrTOUR2-1	AC155344.1	*B. rapa*	285	TAA	GAGACACCCCCATTAGTGAAC	63	Tourist
BrMuMITE5-1	AC155344.1	*B. rapa*	1159	TTTATTaga	354/349	58	Mutator
BoSTOW5-1	AC240081.1	*B. oleracea*	243	TA	TATTTCTTCCGTTTCGATTTA	80	Stowaway
BoTOUR4-1	AC240081.1	*B. oleracea*	267	TAA*	TACTCACTCTGTTTCATAAATGT-CATTCTAACTTTTTT	76	Tourist
Comparison of *B. rapa* BAC AC155341.2 against *B. oleracea* BAC AC240089.1							
BoSTOW4-1	AC240089.1	*B. oleracea*	227	TA	CTGTTTCCGTTTTACAAAGATATACTTTTT	81	Stowaway
b. Identified by presence of TIRs in a single BAC							
BrMuMITE3-1	AC232530.1	*B. rapa*	1586	CAAAAAAAAc	717/689	77	Mutator
c. Identified in *Brassica* nucleotide collection (nr/nt) database by BLASTN searches with reference sequences in Table 2a, b							
BrSTOW1-2	AC232537.1	*B. rapa*	329	TA	GACTCAGGGCCGGCTTACAA	68	Stowaway
BrSTOW1-3	AC232534.1	*B. rapa*	329	TA	GACTCAGGGCCGGCTTACAA	68	Stowaway
BrSTOW1-4	AC189530.2	*B. rapa*	328	TA	GACTCAGGGCCGGCTTACAA	68	Stowaway
BrSTOW1-5	AC189319.1	*B. rapa*	324	TA	G—CAGGGCCGGCT–CAA	68	Stowaway
BoSTOW2-2	EU579455.1	*B. oleracea*	460	TA	GGTGCTAGTCG*	70	Stowaway
BoSTOW2-3	AC152123.1	*B. oleracea*	442	TA	GGCGCTAGTCG*	68	Stowaway
BoSTOW2-4	AC183493.1	*B. oleracea*	436	TA	GGCGCTAGTCG*	72	Stowaway
BrSTOW2-5	AC189511.1	*B. rapa*	422	TA	GGCACTAGTCG*	73	Stowaway
BoSTOW3-2	AC232493.1	*B. oleracea*	244	TA	TGAGAGCATCTTT	66	Stowaway
BoSTOW3-3	AC229603.1	*B. oleracea*	243	TA	GAGCATCTTTAAATA*	58	Stowaway
BoSTOW4-2	AB180902.1	*B. oleracea*	248	TA	CTCCCTCCGTTCGTTAATGATAGAATTTT-TAG	78	Stowaway
BrSTOW4-3	AC189452.2	*B. rapa*	256	TA	CTCTCTCCGTTTCGAAAAGATATATATTT-TAG	82	Stowaway
BrSTOW4-4	AC189417.2	*B. rapa*	253	TA	CTCCTTCCATTTCAAAAAGATAGACTTTT-TAGTA	81	Stowaway
BrSTOW4-5	AC189322.2	*B. rapa*	251	TA	CTCCTTCCGTTTCACAAAGATAGACTTTT-TAG	80	Stowaway
BrSTOW4-6	AC189444.2	*B. rapa*	251	TA	CTCCTTCCGTTCCTAAAATATATACTTTT-TAG	80	Stowaway
BrSTOW4-7	AC232543.1	*B. rapa*	248	TA	CTCCATCCGTCCTAAAAGATAAATTTTT-TAG	79	Stowaway
BrSTOW4-8	AC232514.1	*B. rapa*	245	TA*	CTCCATCCGTTTAAAAAAGATAGATGTTTT	79	Stowaway
BrSTOW4-9	AC189476.2	*B. rapa*	233	TA	CTCTGTTCTTTAAAAATAGATTTTCTAG	79	Stowaway
BrSTOW4-10	AC189492.2	*B. rapa*	218	TA*	CTCCATTCAACAAAAATATATATTTTA	82	Stowaway
BoSTOW5-2	AC240087.1	*B. oleracea*	243	TA	CTCCCTCCGTTTCATATCA	74	Stowaway

Table 2 continued

Element name	BAC Accession	Species	Size	TSDs	Terminal inverted repeats (TIRs)	AT %	Group
BoSTOW5-3	AC183492.1	*B. oleracea*	241	TA	CTCCATCCGTTTCATATTA	74	Stowaway
BrSTOW5-4	AC232467.1	*B. rapa*	244	TA	CTCCCTCCGTTTCGATTTA	76	Stowaway
BrSTOW5-5	AC189391.2	*B. rapa*	242	TA	CTCTCTCCGTTTCATTTTA	74	Stowaway
BnSTOW5-6	AJ291500.1	*B. napus*	242	TA	CTCCCTCTGTTTCATCATA	74	Stowaway
BrSTOW5-7	AC241048.1	*B. rapa*	241	TA	TTCCTTCCGTTTCATTTTA	76	Stowaway
BrSTOW5-8	AC189207.2	*B. rapa*	239	TA*	CTCTCTCCGTTTCATTTTA	78	Stowaway
BrSTOW5-9	AC189417.2	*B. rapa*	242	TA	CTCCCTCCATTTCATTTTA	72	Stowaway
BrSTOW5-10	AC189565.2	*B. rapa*	245	TA	CTCCCTCCATTTTATAATA	78	Stowaway
BrTOUR1-2	AC232445.1	*B. rapa*	421	TAA	GGGGGTGTTAGTG	79	Tourist
BrTOUR1-3	AC189390.2	*B. rapa*	418	TAA	GGGTGTTAGTGGGA	76	Tourist
BrTOUR1-4	AC189314.1	*B. rapa*	413	TTA	GGAGGGTGTTAGTGGGA	76	Tourist
BrTOUR1-5	AC232479.1	*B. rapa*	412	TTA	GGGGGTGTTAGTGGGGA	74	Tourist
BrTOUR1-6	AC189261.2	*B. rapa*	412	TTA	GGGGGTGTTAGTAGGGA	74	Tourist
BrTOUR1-7	AC189219.1	*B. rapa*	412	TTA	GGGGGTGTTAGTGGG	75	Tourist
BnTOUR1-8	AC236791.1	*B. napus*	412	TAA	GGGGGTGTTAGTGAGGA	74	Tourist
BrTOUR1-9	AC189415.2	*B. rapa*	402	TAA	GGGGGTGTTAGTGGG	74	Tourist
BrTOUR1-10	AC189397.2	*B. rapa*	392	TTA	TGGGATATGGATTTGTAGTGA	75	Tourist
BrTOUR2-2	AC172859.1	*B. rapa*	289	TTA	GAGCATCCCCATTAGTGAAC	62	Tourist
BrTOUR2-3	AC189450.2	*B. rapa*	287	TAA	GAGCACCCCCATTAGTGAAC	62	Tourist
BrTOUR2-4	AC189577.2	*B. rapa*	284	TAA	GAGCACCCCATTAGTAAAC	64	Tourist
BrTOUR2-5	AC232550.1	*B. rapa*	273	TTA	GAGCACCCCCATTAGTGAAC	65	Tourist
BoTOUR3-2	DQ222849.1	*B. oleracea*	258	TAA	GGACATCTCCA—(106)	66	Tourist
BoTOUR3-3	DQ222850.1	*B. oleracea*	258	TTA	GAGCATCTCCA—(106)	66	Tourist
BnTOUR3-4	FJ384103.1	*B. napus*	258	TAA	GAGCATCTCCA—(102)	66	Tourist
BrTOUR3-5	AC189458.2	*B. rapa*	258	TTA	GAGCATCTCCA—(102)	67	Tourist
BrTOUR3-6	AC172875.2	*B. rapa*	258	TTA	GAGCATCTCCA—(102)	66	Tourist
BrTOUR3-7	AC189299.2	*B. rapa*	258	TTA	GGGCATCTCCA—(101)	64	Tourist
BrTOUR3-8	AC189445.2	*B. rapa*	258	TTA	GGGCATCTCCA—(103)	67	Tourist
BrTOUR3-9	AC189370.2	*B. rapa*	258	TAA	GAGCATCTCCA—(102)	66	Tourist
BrTOUR3-10	AC155339.1	*B. rapa*	259	TTA	GAGCATCTCCA—(102)	64	Tourist
BrTOUR4-2	AC189192.2	*B. rapa*	332	TTA	CTCCCTCTTCGTAATTAATTACT	77	Tourist
BrTOUR4-3	AC241150.1	*B. rapa*	272	TAA*	TATACTCTCTCTATTTTATAATAAGTGTCA	79	Tourist
BnTOUR4-4	AF136223.1	*B. napus*	272	TAA	TACTCCATCTGTTTCATATTAAGTGTCATT-GTAACA	79	Tourist
BrTOUR4-5	AC232552.1	*B. rapa*	272	TTA	CTACTCCTTCCGTTTCTGAATAAGTGT-CATTTT	78	Tourist
BrTOUR4-6	AC189299.2	*B. rapa*	271	TAA*	TACCCTCTCCATTTCTGAATAACTGTCA	75	Tourist
BrTOUR4-7	AC189587.2	*B. rapa*	266	TCA*	TACTCCTTCCGTTTCTAAATAACTGTCA	81	Tourist
BrTOUR4-8	AC189218.2	*B. rapa*	264	TAA	TACTCTTTCTGTTTCTAAATAAATAT-CACTTTGAAGTTTTT	79	Tourist
BrTOUR4-9	AC189322.2	*B. rapa*	261	TTA	TACTTCCTCCGTTTCATAAAAAATGTCACT	80	Tourist
BrTOUR4-10	AC189569.2	*B. rapa*	255	TAA	TACTCTCTATATTTTTGAAAAAAATAT-CATTT	81	Tourist
BrMUMITE1-2	AC189475.2	*B. rapa*	569	tTTAATGAA	UD	78	Mutator
BrMUMITE1-3	AC189340.1	*B. rapa*	559	TAAAATGAt	UD	78	Mutator
BrMUMITE1-4	AC232437.1	*B. rapa*	547	TTTACATAA	UD	77	Mutator
BnMUMITE1-5	AC236785.1	*B. napus*	527	TATTTaTTaT	UD	77	Mutator
BrMuMITE2-2	AC189218.2	*B. rapa*	1060	TTATTTAAAT	UD	80	Mutator
BrMuMITE2-3	AC189224.1	*B. rapa*	1055	TATTTTATTG	UD	80	Mutator
BrMuMITE2-4	AC189578.2	*B. rapa*	1052	AACAATATAG	UD	80	Mutator

Table 2 continued

Element name	BAC Accession	Species	Size	TSDs	Terminal inverted repeats (TIRs)	AT %	Group
BrMuMITE2-5	AC155345.1	*B. rapa*	958	TAAAACTGTG	UD	81	Mutator
BrMuMITE3-2	AC189366.2	*B. rapa*	1624	AATAAAATAT	UD	78	Mutator
BrMuMITE3-3	AC232539.1	*B. rapa*	1575	CATAATAATT	UD	77	Mutator
BrMuMITE3-4	AC189401.2	*B. rapa*	1555	GATTTAATAT	UD	77	Mutator
BrMuMITE3-5	AC189580.2	*B. rapa*	1497	TAAAAAGAAC	UD	79	Mutator
BrMuMITE3-6	AC232458.1	*B. rapa*	1581	GATTTTCAAG	UD	77	Mutator
BrMuMITE3-7	AC232562.1	*B. rapa*	1552	AAAACAAAAC	UD	77	Mutator
BoMuMITE3-8	EU642504.1	*B. oleracea*	1539	GATTAGATTC	649/616	78	Mutator
BoMuMITE3-9	EU579455.1	*B. oleracea*	886	TTAAATgTT	255/243	78	Mutator
BoMuMITE4-1	AC149635.1	*B. oleracea*	899	TATATATAT	407/446	73	Mutator
BrMuMITE4-3	AC172877.1	*B. rapa*	886	CTAAAATTA	UD	75	Mutator
BrMuMITE4-4	AC232459.1	*B. rapa*	839	ATTTTTCTTT	UD	75	Mutator
BrMuMITE4-5	AB257127.1	*B. rapa*	820	TTTTTTTAA	UD	77	Mutator
BrMuMITE5-2	AC172882.1	*B. rapa*	1157	TTATTaga	354/348	58	Mutator
BrMuMITE5-3	AENI01009313.1	*B. rapa*	1164	AAgAGAAAT	353/357	58	Mutator
BrXMITE1-2	AENI01000925.1	*B. rapa*	356	CTC*	TTCATTTACGTATCGCGCGACCTGCGAC-CTG	52	BoXMITE
BrXMITE1-3	AC189543.2	*B. rapa*	320	AAT*	GGCCTGTTCCTTACCTGTCTGGC	54	BoXMITE
BrXMITE1-4	AENI01003669.1	*B. rapa*	320	AAT*	GGCCTGTTCCTTACCTGTCT	54	BoXMITE
BrXMITE1-5	AENI01006359.1	*B. rapa*	308	CAT*	TCGTTTACGTATCGTGCGACCTGCGACT	58	BoXMITE

The names of each element, their sizes, TSDs, TIRs and derivation are listed. The asterisks in front and small letters in TSDs or TIR indicate a mismatch at 5′ or 3′ TSDs or TIRs

UD undetermined

Table 3 Estimated copy numbers and AT percentage of *Brassica* MITE families

Family	TSDs	TIR length (bp)	Length	No. in database	No. in A-& C-genomes	Average AT %
BrSTOW1	TA	15–30	324–580	52	990	70
BoSTOW2	TA	11	422–460	20	382	71
BoSTOW3	TA	13–15	237–244	12	230	62
BoSTOW4	TA	27–34	218–256	120	2294	80
BoSTOW5	TA	19–21	239–245	170	3239	76
BrTOUR1	TNA	13–21	392–421	85	1624	75
BrTOUR2	TNA	19–21	273–289	64	1224	63
BrTOUR3	TNA	102–106	258–259	205	3918	66
BoTOUR4	TNA	23–41	255–332	128	2446	78
BrMuMITE1	9 bp	122–125	527–569	256	4892	78
BrMuMITE2	10 bp	400–450	905–1060	22	420	80
BrMuMITE3	10 bp	250–750	1497–1624	28	535	78
BoMuMITE4	9 bp	556–446	766–899	312	5964	75
BrMuMITE5	9 bp	348–357	1152–1164	6	114	58
BoXMITE1	TTC	21–42	308–402	12	229	53

The average lengths of elements and their average AT contents are given

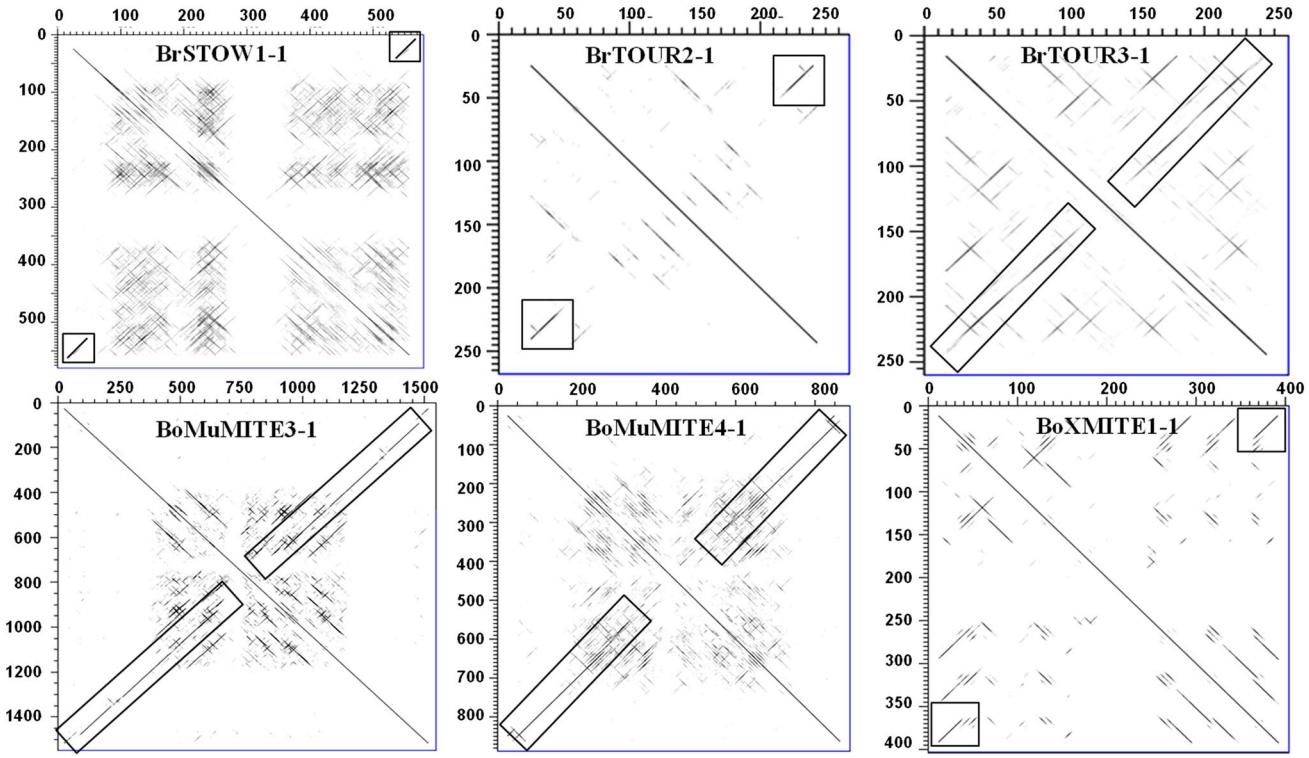

Fig. 1 Structural characterization of MITEs in *Brassica* by *dot plot* sequence analysis. Each *dot plot* of a MITE against itself allowed the identification of TIRs at corners (*boxed*). The *central complete diagonal line* represents self-homology, while the *boxed inverse-diagonal lines* show the TIRs; *other lines* show near-direct repeats in both forward and inverted orientations. *Scales in base pairs* show the wide range in sizes of elements and length of TIRs

in *BoSTOW4* and 5 bp in *BrTOUR3* are low information content (low-height) letters representing non-homologous sequence or an insertion. While showing conservation within each of the 15 families, TIRs varied extensively between most families (Fig. 2). Length differed from 11 to 106 bp, some showed near equal AT-CG content while others were AT-rich or even showed gross strand asymmetry in base pair composition (up to 93 % TG rich in one strand of *BrTOUR1*), and there was a varying degree of conservation of 3′ or 5′ termini or internal domains within individual families. Mutator-derived MITEs showed highly AT-rich regions within TIRs (Table 2) and internal regions.

Site-specific insertion polymorphism of MITEs in *Brassica* germplasm

To investigate the polymorphisms of *Brassica* MITEs among 40 *Brassica* accessions (Table 1) from three diploids (AA, BB, CC), three allotetraploids (AABB, AACC, BBCC) and 2 synthetic hexaploid *Brassicas* (AABBCC, *B. napus* × *B. nigra*; *B. carinata* × *B. rapa*), primer pairs (Table 4) were designed from sequences flanking the MITEs identified by comparison of homoeologous BAC pairs. The insertion polymorphisms (Table 2a) showed that some families had been active after the evolutionary separation of the genomes.

Stowaway MITE insertion polymorphisms

The presence of the 237–244 bp *BrSTOW3-1* elements was tested using primer pair BoSTOW3F/R (Table 4; Fig. 3), with a product size of 512 bp including the MITE element or 272 bp where the element was not present (flanking region with pre-insertion or empty sites). All *B. oleracea* lines included the element at this site, while the insert was absent in *B. rapa* and *B. nigra* accessions. The *BrSTOW3-1* element was also present in the allotetraploids *B. napus* (AACC) and *B. carinata* (BBCC) and four hexaploid *Brassica* lines (AABBCC), but absent in *B. juncea* (AABB), consistent with its presence only in C-genome diploid except in the Pakistani accession line 12 (NARC-II; see "Discussion"). The *BoSTOW4-1* (Fig. 3c) insertion proved to be like *BrSTOW3-1*, specific to the C-genome, with presence of the element indicated by a 500 bp band in all the accessions with a C-genome, and a 273 bp band representing flanking sequences only in the A- and B-genomes.

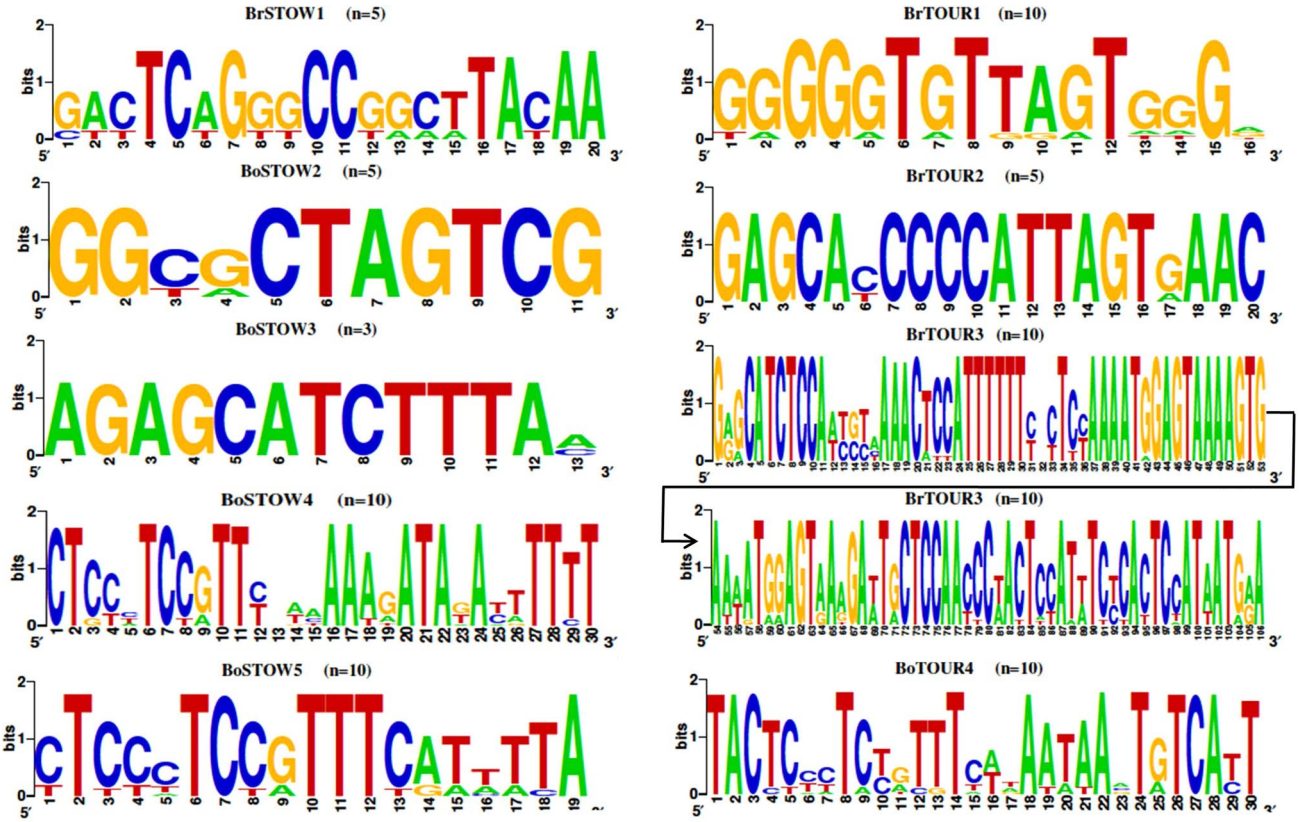

Fig. 2 Sequence logos (pictograms) of *Brassica* MITE TIRs. The logos were generated with (*n*) sequences, and letter heights (0–2 bits) indicate the information content of consensus nucleotides at each position in the TIRs of *Brassica Stowaway* (*left*) and *Tourist* (*right*) MITEs. Lower heights represent non-conserved motifs or insertions within a family. There is little conservation between the families

Table 4 *Brassica* MITE primers with size of the elements, size of the expected products, names and sequence of primers

MITEs derivation	Element size	Product size	Targeted MITE insertion	Primers	Primer sequence
Stowaway	324–580	682	*BrSTOW1-1*	BrSTOW1F BrSTOW1R	CTTCGTATTCTCTGCAAGAT CGAAATACATAGACGTATAC
Stowaway	237–244	512	*BoSTOW3-1*	BoSTOW3F BoSTOW3R	AGGGTCCAAACATGTGATTA GTTGCAAATAATTGATCGTTG
Stowaway	227	500	*BoSTOW4-1*	BoSTOW4F BoSTOW4R	CAATACCATCCAGTGTTACA TGTTGTCGTCATTAAGGTGA
Tourist	392–421	530	*BrTOUR1-1*	BrTOUR1F BrTOUR1R	GGGGATAATTACACATCTTG CAAATCTCCGACATCAATC
Tourist	273–289	510	*BrTOUR2-1*	BrTOUR2F BrTOUR2R	AGGGTCCAAACATGTGATTA GTTGCAAATAATTGATCGTTG
Tourist	258	564	*BrTOUR3-1*	BrTOUR3F BrTOUR3R	GGACCATACAGTATATCGTT TGGATAACGTTGTTGTTCCC
Mutator	527–569	1016	*BrMuMITE1-1*	BrMuMITE1F BrMuMITE1R	CATTGCAGAAGAGCTGGCTGC CAAGATTTTGAGGAGAGATTTG
Mutator	766–899	990	*BrMuMITE4-2*	BrMuMITE4F BrMuMITE4R	GATAATTTTTGGGCCATGCA CGATCAGACAAACGACGAAA

The organization of the element and flanking sequences for *BrSTOW1-1* was more complex (Fig. 3a), with only one (cultivar De Rosny) accession of *B. oleracea* (CC) and *B. carinata* (BBCC) showing any amplification using the BrSTOW1F/R primer pair, suggesting divergence or loss of the flanking sites in the B- and C-genome ancestor compared to the A-genome. Of species including the A-genome, some showed sites without the inserted

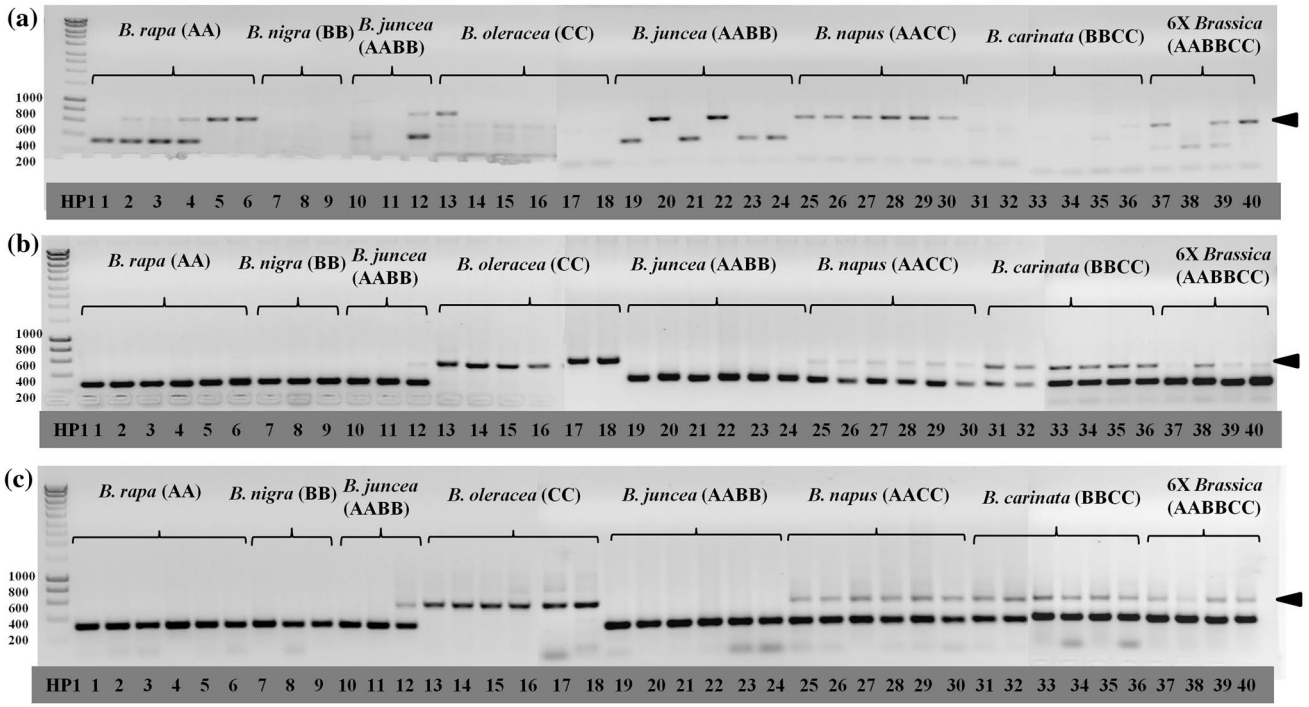

Fig. 3 Insertion polymorphisms of *Brassica* Stowaway-like MITEs amplified by PCR with flanking primers. **a** *BrSTOW1-1* and **b** *BoSTOW3-1* c *BoSTOW4-1*. Figures 3–5 show inverted images of ethidium bromide stained PCR-amplified DNA after size separation by agarose gel electrophoresis. Lower numbers (1–40) identify individual lanes for each *Brassica* accession listed in Table 1. Braces group *Brassica* species. *Black arrowheads* (*right*) *upper bands* with amplified loci having MITE insertions while *lower bands* amplify the loci without insertions. *Left lane* (HP1) is 200 bp marker ladder (Hyperladder I) with band sizes indicated

element, some showed presence of the Stowaway element (682 bp), and another group, including three of the diploid *B. rapa*, showed bands associated with both presence and absence of the Stowaway element; surprising in these inbred lines expected to be homozygous. Therefore, it seems that the region flanking the Stowaway element is duplicated in the genome, allowing amplification of sites with and without its insertion, a duplication that is not shared by only one of the 15 A-genome tetraploids. A few faint bands were interpreted as amplification between sites with weak homology to the primers.

Tourist MITE insertion polymorphisms

The primer pair BrTOUR1F/R amplified *BrTOUR1-1* products with the MITE insertion from 15 of the 40 *Brassica* genomic DNA accessions, as well as a shorter primer-related product from all 40 accessions (Fig. 4a). It did not amplify MITEs in the three B-genome accessions, but amplified only from one A-genome accession (cultivar Chinese Wong Bok). Of the six diploid C-genomes, five showed amplification of the MITE elements. Consistent with these results, the AB-genome *B. juncea* accessions showed no amplification except for the anomalous accession (NARC-II; see "Discussion"), and five of the six *B. napus* accessions showed amplification. Despite including the C-genome, there was no amplification from the six *B. carinata* accessions.

The primers BrTOUR2F/R amplified a 510 bp product with the MITE insertions in accessions with the A-genomes including all *B. rapa* and *B. juncea* lines, six lines from *B. napus* and all four hexaploid *Brassica* lines (Fig. 4b). Of accessions with only B- and/or C-genomes, only one *B. carinata* (NARC-PK; Pakistani origin) showed amplification of the MITE element. The primers for *BrTOUR3-1* did not detect the element in the B- and C-genomes (306 bp band). Bands including the length of the *BrTOUR3-1* MITE insertions (564 bp) were found in *B. rapa* (A) and the tetraploids AABB and AACC *Brassicas* (Fig. 4c). Mutation of a primer site (or, less likely, evolutionary mobility of the element) may lead to results showing two bands (with and without the element) in four of the six *B. rapa* accessions, and the absence of lower bands in four of the nine *B. juncea* accessions (which would be expected to include the empty site from the B-genome).

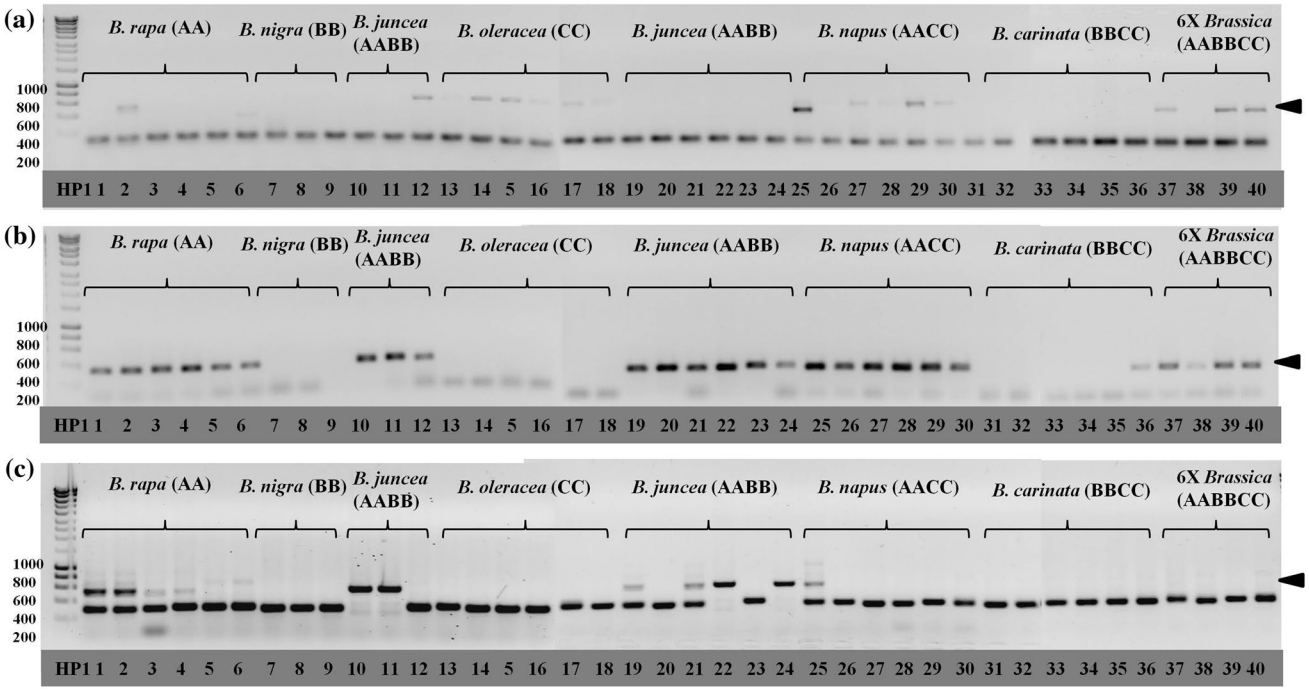

Fig. 4 Insertion polymorphism of *Brassica* Tourist-like MITEs amplified by PCR with flanking primers. **a** *BrTOUR1-1*, **b** *BrTOUR2-1* and **c** *BrTOUR3-1*. See Fig. 3 for explanation

Mutator MITE insertion polymorphisms

The primers for *BrMuMITE1-1* amplified 1016 bp sites where the *Mutator*-derived MITE was present (Fig. 5a). The primers showed some homology to other sites in the genome and hence weak amplification of other products. One *B. rapa* accession (San Yue Man) showed a major product corresponding to the 508 bp of the empty site seen in the source BAC sequence, along with a product corresponding to the insertion and an intermediate product, suggesting genomic rearrangement and duplication associated with the MITE element. No strong amplification was seen in any B- or C-genome accessions, and all the tetraploids with the A-genome showed presence of the MITE. *BoMuMITE4-2* (Fig. 5b) showed many polymorphisms between accessions and species, with null sites (no product), amplification across empty sites (400 bp product), and amplification with the *BoMuMITE4* insertion (990 bp product).

Chromosomes and genomic localization of MITEs in *Brassica*

We studied the localization and distribution of high-copy-number (Table 3) MITE elements on *Brassica* chromosomes. The Mutator-derived MITEs were amplified and labelled, before hybridization to chromosomes from the allotetraploid (4*x*) *Brassica* species. *BrMuMITE1-1* showed dispersed hybridization to the 20 A-genome chromosomes in *B. napus* (AACC, $2n = 38$; Fig. 6a, b) and *B. juncea* (AABB, $2n = 36$; Fig. 6e). In contrast, *BoMuMITE4-2* was most abundant in sub-telomeric regions of particular chromosomes from both genomes, with much weaker dispersed hybridization on A-genome chromosomes. In both *B. napus* (Fig. 6b, c) and *B. carinata* (BBCC, $2n = 34$; Fig. 6f), *BoMuMITE4-2* showed colocalization with major and minor rDNA sites (constrictions and weaker DAPI staining). Chromosome number in *B. juncea* (AABB, $2n = 4x = 36$) line NARC-II was confirmed by DAPI staining (Fig. 6d), as PCR results from this accession were anomalous (see above and "Discussion").

Structural features of an unknown MITE family in *Brassica*

We identified a MITE-like element with 3 bp TSDs (nucleotide sequence TTC) and 42 bp imperfect TIRs but no significant homology to any known MITEs. The element is named *BoXMITE1-1* and represents a low-copy-number family (*BoXMITE1*) with only 229 estimated copies within *Brassica* (A-, C-genomes, Table 3). *BoXMITE1-1*, the first identified element from the family was found inserted in *B. oleracea* BAC sequence accession EU642504.1 from 86275 to 86676 bp. Using this as query sequence against the *Brassica* nucleotide collection database in GenBank, only two complete sequences were retrieved,

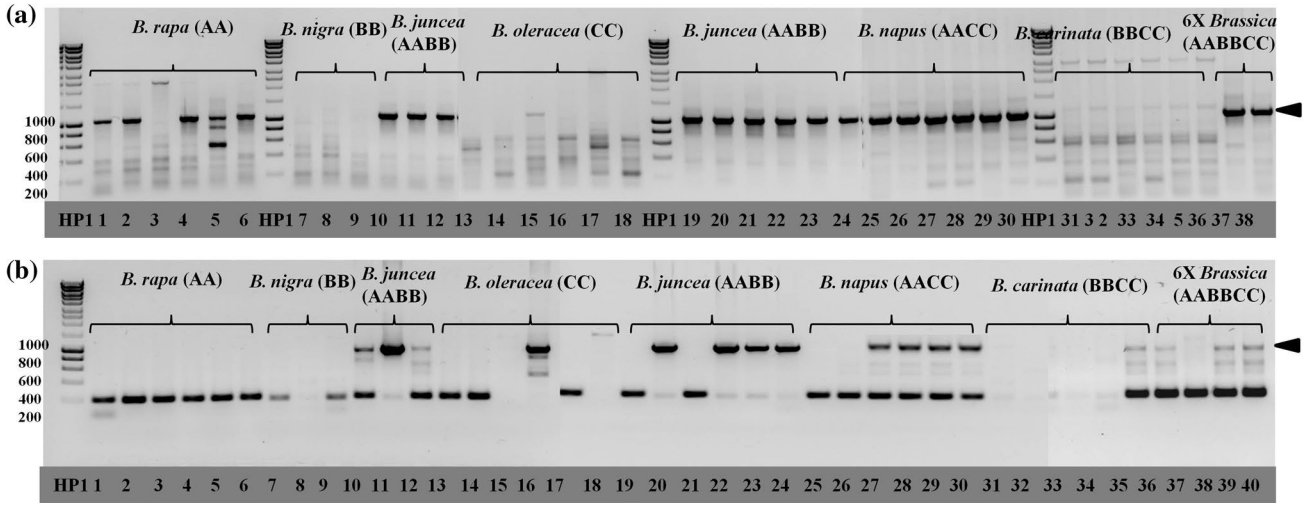

Fig. 5 Insertion polymorphism of *Brassica* Mutator-like MITEs amplified by PCR with flanking primers. **a** *BrMuMITE1-1* and **b** *BoMuMITE4-2*. See Fig. 3 for explanation

while searching against *Brassica* whole genome shotgun contigs (wgs) database, we collected another 3 full length copies (*BrXMITE1-2, BrXMITE1-4, BrXMITE1-5*). The annotations indicate their localization on chromosome 1, 4 and 7 of *B. rapa*. The elements of the family range in size from ~308 to 402 bp with 3 bp TSDs with single bp mismatch at variable positions. The TIRs of the family members range from 21 to 42 bp with few bp mismatches. *BoXMITE1-1* is flanked by 42 bp, while *BrXMITE1-4* is flanked by 21 bp TIRs (Table 3).

Discussion

Our molecular characterization of 15 novel MITE families (Tables 2, 3) from *Brassica* showed five were derived from Stowaway-like elements, four were Tourist-like, five Mutator-like and one is a novel MITE family (*BoXMITE1*), whose progenitors were not identified. Except for *BrMuMITE3*, the first reference member of all these families was identified by dot plot comparison of homologous BAC sequences of diploid *Brassica* genomes, indicating MITE presence in one species and absence in the other. This strategy identifies elements which have shown mobility since evolutionary separation of the diploid species from a common ancestor, and is not dependent on previous knowledge or identification of sequences related to reference elements. The abundant, high-copy-number elements showed structural characteristics of MITEs (Figs. 1, 2) and TSDs related to the known Stowaway, Tourist and Mutator groups (Tables 2, 3). One novel group, *BoXMITE1*, had a lower copy number with unusual TIR and TSD structures consistent with its MITE origin. It is notable that all the MITE families show activity since the separation of the species.

Approximately 30,000 MITE-like sequences belonging to the 15 families were estimated to occur in the *B. rapa* and *B. oleracea* genomes (Table 3). Around 45,821 MITE sequences belonging to 174 families were identified in *B. rapa* using MITE Digger and MITES-Hunter programs (Chen et al. 2013). *BraSto*, a well-characterized Stowaway MITE family was reported with similar abundance to our MITE families in *Brassica* (Sarilar et al. 2011). The rice genome harbours rather more elements, with ~178,533 MITE-related sequences clustering into 338 families (Lu et al. 2012). A parallel study in the *Solanaceae* has revealed a high level of MITE diversity among the crops in the family (tomato, potato and tobacco) and 22 families including derivatives of Stowaway, Tourist, hAT, and Mutator-like MITEs (Kuang et al. 2009). Several CACTA and hAT-like non-autonomous families were also investigated from *Brassica* (not discussed here), while hAT-derived MITEs have been studied independently in various species, e.g. in *Beta vulgaris* and *Musa* (Menzel et al. 2012, 2014). Given the activity and polymorphisms of the elements and the presence of LTRs, whole genome assembly approaches (often published without details of parameters) may well collapse or delete MITEs so copy number estimates from whole sequence data may be wildly inaccurate.

Evolution and biodiversity of MITEs: amplification and insertion polymorphisms

MITEs transpose into new sites, with or without replication at variable rates (influenced by genomic stress and

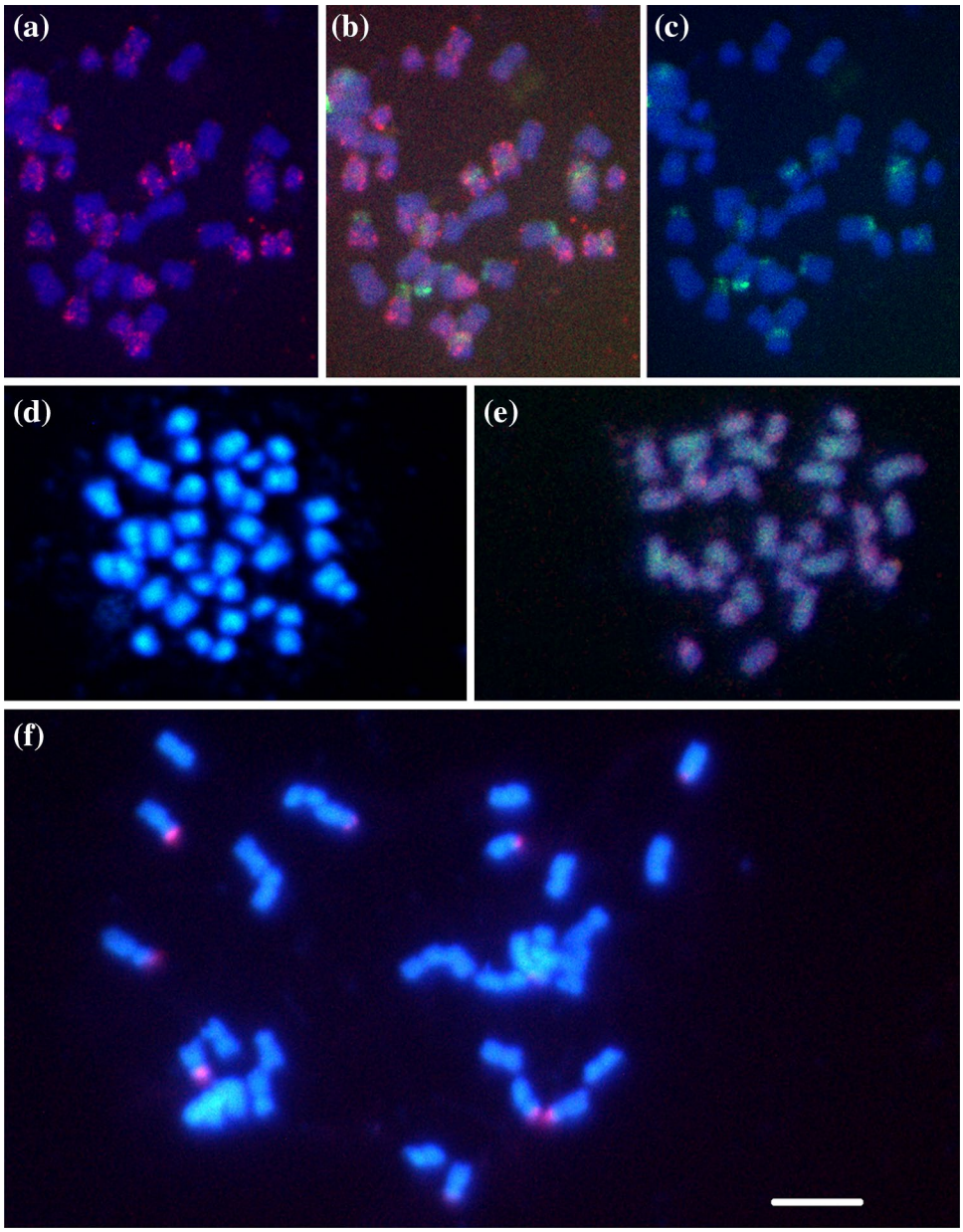

Fig. 6 Fluorescent in situ hybridization (*green* and *red* signals) showing locations of MITEs on *Brassica* metaphase chromosomes stained with DAPI (*blue*). **a**–**c** *B. napus* ($2n = 4x = 38$ AACC) with (**a**, **b**) *BrMuMITE1-1* (*red*) labelling the 20 A-genome chromosomes along most of their length with some stronger sites, and (**b**, **c**) *BoMuMITE4-2* (*green*) labelling about 14 sites near 45S rDNA loci and some dispersed signal primarily on A-genome chromosomes. **d** A metaphase of *B. juncea* line NARC-II (PK001325; $2n = 4x = 36$ AABB) stained with DAPI; many PCR results from this accession were anomalous. **e** *BrMuMITE1-1* (*red*) hybridized to a metaphase of *B. juncea* showing strong hybridization to A-genome chromosomes (excluding some centromeric regions) and very weak hybridization to chromosomes of B-genome origin. **f** Metaphase chromosomes of *B. carinata* ($2n = 4x = 34$ BBCC) showing *BoMuMITE4-2* (*red*). *Scale bar* 8 μm (color figure online)

hybridization; Madlung and Comai 2004) in different cultivars or genotypes, creating the presence/absence-based polymorphisms (Lyons et al. 2008) as described for retrotransposons in the RBIP analysis (Flavell et al. 1998) and exploited here to identify non-selectively all the MITE families in *Brassica* (Fig. 1). Compared to a site containing a MITE, an 'empty site' may be detected either where no element has been in the genomic sequence, or after a MITE excises and moves, when the empty donor host site exhibits a footprint usually with an extra TSD sequence compared to the locus prior to MITE insertion.

Individual transposons, including MITEs, differ in their conservation and proliferation properties (Kubis et al. 1998; Feschotte and Mouches 2000). High conservation in a genome can indicate recent amplification as a burst, while presence over a wide evolutionary lineage shows ancient amplification (Oki et al. 2008; Zerjal et al. 2012). We exploited the knowledge that the MITEs identified here are evolutionarily active to characterize their presence in diverse *Brassica* germplasm, and reconstruct lineages. Primers flanking the MITEs were designed from genomic DNA conserved between two species for PCR of genomic DNA, and the insertion polymorphism of *Brassica* Stowaway, Tourist and Mutator-derived MITEs was observed among 40 cultivars (Figs. 3, 4, 5). The PCR amplification of *BrSTOW1-1* in *B. rapa* and *BoSTOW3-1* and *BoSTOW4-1* in *B. oleracea* suggested the conservation of MITEs in A- and C-genomes, with empty sites (shorter

products) observed in some lineages (Fig. 3). Similarly, the amplification of *Brassica* Tourist and Mutator MITEs (Figs. 4, 5) yielded products with and without insert, displaying insertion polymorphisms. The polymorphisms of particular elements enabled identification and differentiation of many cultivars in *Brassica*; MITE-related molecular markers were used in other plants such as barley (Lyons et al. 2008) and maize (Lu et al. 2012) to study the biodiversity and evolutionary phenomena.

The high-copy-number families (Table 3) related to the two individual MuMITE elements used for PCR amplification in Fig. 5 were used for in situ hybridization to *Brassica* chromosomes (Fig. 6). The genomic location of the individual elements is unknown, but the families showed contrasting distributions: *BrMuMITE1-1* was amplified and dispersed overall A-genome chromosomes. *BoMuMITE4-2* was present on both genomes, co-localizing with 45S rDNA sites (despite the AT richness of MITE sequences, while rDNA sequences are GC rich, seen by their weaker DAPI staining on chromosome preparations), and also showed weaker, dispersed hybridization, greater on the C-genome chromosomes. The proliferation and genomic locations of the families (Fig. 6) are consistent with copy number estimates (Table 3), and PCR amplification results of the single family members (Fig. 5), with a contrast between *BrMuMITE1-1* (isolated from *B. rapa*), being largely A-genome-specific, and *BoMuMITE4-2* (from *B. oleracea*) present in the C-genome but being more polymorphic, less genome-specific and apparently targeted to rDNA sites. While transposon association with rDNA is unusual (and indeed, retrotransposons are often excluded from rDNA loci, e.g. Brandes et al. 1997; Kuipers et al. 1998), rDNA-associated SINE localization is reported in *Brassica* (Goubely et al. 1999). Recently, Eagle and Crease (2012) have reported a DNA transposon associated with complex amplification and rearrangement events in rDNA loci in *Daphnia* that, like *BoMuMITE4*, targets rDNA and also occurs in other genomic locations.

Origins and genomic constitution of *Brassica* accessions

The PCR insertion polymorphism (Figs. 3, 4, 5) gave results that were generally consistent with the presence or absence of polymorphisms in the diploid species and genome constitutions of the tetraploid species. *BrSTOW1-1*, *BrTOUR1-1*, and *BrTOUR3-1* were polymorphic in diploid genomes, also reflected in the tetraploids, thus supporting their polyphyletic origin (Cifuentes et al. 2010). However, a few accessions analysed here showed results that were not consistent with their morphological identification. In particular, accession NARC-II (line 12; PK-001325) from Pakistan, morphologically and by chromosome number (previously unconfirmed, Fig. 6d) a *B. juncea* (AABB) accession, showed properties of the presence of the C-genome as well as A- and B-genomes. Other lines from Pakistan, including *B. carinata* accession NARC-PK (line 36; PK-0085490) (*MuMITE4-2*, *BoTOUR2-1*), showed some results that were not consistent with their expected genomic origin. It is notable that there is a long history of interspecific hybridization and intercrossing of *Brassica* species within Punjab region (Pakistan and India), where these accessions originate (Sikka 1940). It is therefore possible that current accessions have a hybrid ancestry: accession NARC-II (line 12; PK-001325) shows strong evidence for the presence of MITE elements from all three A-, B- and C-genomes. It will be valuable to characterize further these accessions using more genome-specific markers or genome-specific probes for fluorescent in situ hybridization. The exploitation of agronomic and quality characters introgressed from different *Brassica* genomes is an important target for breeders (Tu et al. 2009; Kumar et al. 2011; Heneen et al. 2012).

In *B. juncea*, *BrTOUR3-1* shows accessions that are identical ('homozygous') at all four sites both for presence and absence of the insertion, and with both alleles. It is unexpected that the empty site from the B-genome diploids is not seen in four of the nine *B. juncea* AB tetraploids (Fig. 4). Both *BrSTOW1-1* and *BrTOUR1-1* unexpectedly have heterozygosity in the inbred A-genome diploid (as well as in the tetraploids; Figs. 3, 4); given that these are inbred *Brassica* lines; it is possible that there is duplication of flanking sequences (also possible for the lower band amplified with *BrTOUR1-1* primers). As more sequence for the B-genome is obtained, it will be important to identify elements with specificity to this lineage.

Brassica MITEs display high AT-rich regions

One of the typical features of MITEs is the presence of highly AT-rich sequences (e.g. the *AhMITEs* from *Arachis hypogea* exhibit an AT content of 70 %; Shirasawa et al. 2012), a characteristic found in all *Brassica* MITEs. The average AT contents (Table 3) within the *Brassica* MITE families range from 53 % (*BoXMITE1*) to 80 % (*BoSTOW4*, *BrMuMITE2*).

Conclusions

Our results show that truncated derivatives of various autonomous DNA transposons superfamilies designated as MITEs, detected by bioinformatics and molecular techniques, are evolutionarily active and dispersed in *Brassica* genomes, and some have shown polymorphisms in different genotypes. Thus, MITEs are playing a role in diversification and evolution of the *Brassica* genome. The present

work identifies the range of MITE families in *Brassica* and enables their identification, characterization and annotation as well as study of distribution, diversity and mobility. The study of their flanking genomic sequences and insertion polymorphisms, consequent on their transposition activity, suggests that MITE mobility played an important role in mechanism of genome evolution and diversification. MITEs have potential use as gene modifiers or mutagens. The identification of *Brassica* MITEs will have broad applications in *Brassica* genomics, breeding, hybridization and phylogeny through their use as DNA markers.

Acknowledgments This work was supported by Post-quake Development Plan, Hazara University Mansehra, Pakistan. We are grateful to Hazara University and Higher Education Commission of Pakistan for funding and support of this work. We thank Dr. Graham Teakle and Dr. Guy Barker from Warwick University, UK and Dr. Xian Hong Ge from University of Wuhan, China for supplying seeds or DNA from the *Brassica* accessions studied.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Altschul SF, Gertz EM, Agarwala R, Schaffer AA, Yu YK (2009) PSI-BLAST pseudocounts and the minimum description length principle. Nucleic Acids Res 37:815–824. doi:10.1093/nar/gkn981

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. doi:10.1093/gbe/evp009

Bergemann M, Lespinet O, M'Barek SB, Daboussi MJ, Dufresne M (2008) Genome-wide analysis of the *Fusarium oxysporum* mimp family of MITEs and mobilization of both native and de novo created mimps. J Mol Evol 67(6):631–642. doi:10.1007/s00239-008-9164-7

Brandes A, Heslop-Harrison JS, Kamm A, Kubis T, Doudrick T, Schmidt T (1997) Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol Biol 33:11–21

Bureau TE, Wessler SR (1992) Tourist: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283–1294. doi:10.1105/tpc.4.10.1283

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 6:907–916. doi:10.1105/tpc.6.6.907

Capy P (2005) Classification and nomenclature of retrotransposable elements. Cytogenet Genome Res 110:457–461. doi:10.1159/000084978

Chen J, Hu Q, Zhang Y, Lu C, Kuang H (2013) P-MITE: a database for plant miniature inverted-repeat transposable elements. Nucleic Acid Res: 1–6. doi:10.1093/nar/gkt1000

Cifuentes M, Eber F, Lucas MO, Lode M, Chevre AM, Jenczewski E (2010) Repeated polyploidy drove different levels of crossover suppression between homoeologous chromosomes in *Brassica napus* allohaploids. Plant Cell 22(7):2265–2276. doi:10.1105/tpc.109.072991

Eagle S, Crease, T (2012) Copy number variation of ribosomal DNA and Pokey transposons in natural populations of *Daphnia*. Mobile DNA 3(1). URL http://dx.doi.org/10.1186/1759-8753-3-4

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Feschotte C, Mouches C (2000) Evidence that a family of miniature inverted-repeat transposable elements (MITEs) from the *Arabidopsis thaliana* genome has arisen from a pogo-like DNA transposon. Mol Biol Evol 17:730–737

Flavell AJ, Knox MR, Pearce SR, Ellis TH (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16:643–650

Ge XH, Wang J, Li ZY (2009) Different genome-specific chromosome stabilities in synthetic *Brassica* allohexaploids revealed by wide crosses with *Orychophragmus*. Ann Botany 104:19–31. doi:10.1093/aob/mcp099

Goubely C, Arnaud P, Tatout C, Heslop-Harrison JS, Deragon JM (1999) S1 SINE retroposons are methylated at symmetrical and non-symmetrical positions in *Brassica napus*: identification of a preferred target site for asymmetrical methylation. Plant Mol Biol 39:243–255

Heneen WK, Geleta M, Brismar K, Xiong Z, Pires JC, Hasterok R, Stoute AJ, Scott RJ, King GJ, Kurup S (2012) Seed colour loci, homoeology and linkage groups of the C genome chromosomes revealed in *Brassica rapa–B. oleracea* monosomic alien addition lines. Ann Botany 109(7):1227–1242. doi:10.1093/aob/mcs052

Hikosaka A, Nishimura K, Hikosaka-Katayama T, Kawahara A (2011) Recent transposition activity of *Xenopus* T2 family miniature inverted-repeat transposable elements. Mol Genet Genomics 285(3):219–224. doi:10.1007/s00438-010-0599-3

Jiang N, Bao Z, Zhang X, Eddy SR, Wessler SR (2004a) Pack-MULE transposable elements mediate gene evolution in plants. Nature 431:569–573. doi:10.1038/nature02953

Jiang N, Feschotte C, Zhang X, Wessler SR (2004b) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr Opin Plant Biol 7:115–119. doi:10.1016/j.pbi.2004.01.004

Junier T, Pagni M (2000) Dotlet: diagonal plots in a web browser. Bioinformatics 16:178–179

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110:462–467. doi:10.1159/000084979

Kuang H, Padmanabhan C, Li F, Kamei A, Bhaskar PB, Ouyang S, Jiang J, Buell CR, Baker B (2009) Identification of miniature inverted-repeat transposable elements (MITEs) and biogenesis of their siRNAs in the *Solanaceae*: new functional implications for MITEs. Genome Res 19:42–56. doi:10.1101/gr.078196.108

Kubis SE, Heslop-Harrison JS, Desel C, Schmidt T (1998) The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. Plant Mol Biol 36: 821–831. http://aob.oxfordjournals.org/content/82/suppl_1/45.abstract

Kuhn GC, Heslop-Harrison JS (2011) Characterization and genomic organization of PERI, a repetitive DNA in the *Drosophila buzzatii* cluster related to DINE-1 transposable elements and highly abundant in the sex chromosomes. Cytogenet Genome Res 132(1–2):79–88. doi:10.1159/000320921

Kuipers AG, Heslop-Harrison JS, Jacobsen E (1998) Characterisation and physical localisation of Ty1-copia-like retrotransposons in four *Alstroemeria* species. Genome 41(3):357–367

Kumar S, Atri C, Sangha MK, Banga S (2011) Screening of wild crucifers for resistance to mustard aphid, *Lipaphis erysimi* (kaltenbach) and attempt at introgression of resistance gene(s) from *Brassica fruticulosa* to *Brassica juncea*. Euphytica 179(3):461–470

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 29:1005–1017. doi:10.1093/molbev/msr282

Lyons M, Cardle L, Rostoks N, Waugh R, Flavell AJ (2008) Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome. Mol Genet Genomics 280:275–285. doi:10.1007/s00438-008-0363-0

Madlung A, Comai L (2004). The effect of stress on genome regulation and structure. Ann Botany 94 (4): 481–495. URL http://dx.doi.org/10.1093/aob/mch172

Menzel G, Krebs C, Diez M, Holtgräwe D, Weisshaar B, Minoche AE, Dohm JC, Himmelbauer H, Schmidt T (2012) Survey of sugar beet (*Beta vulgaris* L.) hAT transposons and MITE-like hATpin derivatives. Plant Mol Biol 78:393–405. doi:10.1007/s11103-011-9872-z

Menzel G, Heitkam T, Seibt KM, Nouroz F, Müller-Stoerme M, HeslopHarrison JS, Schmidt T (2014) The diversification and activity of hAT transposon in *Musa* genome. Chr Res 22(4):559–571

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. Gen Genetic Sys 83:321–329

Ouyang S, Buell CR (2004) The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Res 32:D360–D363. doi:10.1093/nar/gkh099

Sarilar V, Marmagne A, Brabant P, Joets J, Alix K (2011) BraSto, a Stowaway MITE from *Brassica*: recently active copies preferentially accumulate in the gene space. Plant Mol Biol 77:59–75. doi:10.1007/s11103-011-9794-9

Shirasawa K, Hirakawa H, Tabata S, Hasegawa M, Kiyoshima H, Suzuki S, Sasamoto S, Watanabe A, Fujishiro T, Isobe S (2012) Characterization of active miniature inverted-repeat transposable elements in peanut genome. Theor Appl Genet 124(8):1429–1438. doi:10.1007/s00122-012-1798-6

Sikka SM (1940) Cytogenetics of *Brassica* hybrids and species. J Genetics 40:441–509

Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167: GC1-10

Sun G, Pourkheirandish M, Komatsuda T (2009) Molecular evolution and phylogeny of the RPB2 gene in the genus *Hordeum*. Ann Botany 103 (6): 975–983. URL http://dx.doi.org/10.1093/aob/mcp020

Tu Z (2001) Eight novel families of miniature inverted repeat transposable elements in the African malaria mosquito, *Anopheles gambiae*. Proc Natl Acad Sci USA 98:1699–1704. doi:10.1073/pnas.041593198

Tu Y, Sun J, Ge X, Li Z (2009) Chromosome elimination, addition and introgression in intertribal partial hybrids between *Brassica rapa* and *Isatis indigotica*. Ann Botany 103 (7): 1039–1048. URL http://dx.doi.org/10.1093/aob/mcp045

Walley PG, Teakle GR, Moore JD, Allender CJ, Pink DAC, Buchanan-Wollaston V, Barker G (2012) Developing genetic resources for pre-breeding in *Brassica oleracea* l: an overview of the UK perspective. J Plant Biol 39(1):62–68

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genetics 5:814–821

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Gen 8:973–982. doi:10.1038/nrg2165

Yang G, Hall TC (2003) MAK, a computational tool kit for automated MITE analysis. Nucleic Acids Res 31:3659–3665

Zerjal T, Rousselet A, Mhiri C, Combes V, Madur D, Grandbastien MA, Charcosset A, Tenaillon MI (2012) Maize genetic diversity and association mapping using transposable element insertion polymorphisms. Theor Appl Genet 124(8):1521–1537. doi:10.1007/s00122-012-1807-9

Zhang X, Jiang N, Feschotte C, Wessler SR (2004) PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. Genetics 166:971–986