Pak. J. Bot., 47(4): 1367-1375, 2015.

MOLECULAR CHARACTERIZATION AND DIVERSITY OF A NOVEL NON-AUTONOMOUS *MUTATOR-LIKE* TRANSPOSON FAMILY IN *BRASSICA*

FAISAL NOUROZ[1,2], SHUMAILA NOREEN[3] AND J.S. HESLOP-HARRISON[2]

[1]*Department of Botany, Hazara University Mansehra Pakistan*
[2]*Department of Biology, University of Leicester, Leicester LE1 7RH, UK*
[3]*Department of Genetics, University of Leicester, LE1 7RH, UK*
**Corresponding author e-mail: faisalnouroz@gmail.com, +92 3355851620*

Abstract

Transposable elements (TEs) are capable of mobilizing from one genomic location to other, with changes in their copy numbers. *Mutator-like* elements (*MULEs*) are DNA transposons characterized by 9 bp target site duplications (TSDs), with high variability in sequence and length, and include non-conserved terminal inverted repeats (TIRs). We identified and characterized a family of *Mutator-like* elements designated as *Shahroz*. The structural and molecular analyses revealed that family had a small number of mostly defective non-autonomous *MULEs* and has shown limited activity in the evolutionary history of the *Brassica* A-genome. The *Shahroz* elements range in size from 2734 to 3160 bp including 76 bp imperfect TIRs and 9 bp variable TSDs. The individual copies have shown high homology (52–99%) in their entire lengths. The study revealed that the elements are less in numbers but active in *Brassica rapa* genomes and PCR amplification revealed their specificity and amplification in A-genome containing diploid and polyploids *Brassica*. The phylogenetic analysis of *Brassica MULEs* with other plant *Mutator* elements revealed that no correlation exists between *Brassica MULEs* and other elements suggesting a separate line of evolution. Analyzing the regions flanking the insertions revealed that the insertions have showed a preference for AT rich regions. The detailed study of these insertions revealed that although less in number and small sizes, they have played a role in *Brassica* genome evolution by their mobilization.

Key words: Transposable elements, *Brassica*, Mutator, Diversity, Shahroz, Phylogenetic analysis.

Introduction

Transposable elements (TEs), mobile genetic elements are a major component of all eukaryotic genomes, representing 40% of the entire genome in humans (Mills *et al*., 2006) and 50-90% of the genome in important agricultural crops like maize, wheat, barley, rye, sugar beet or *Vicia* (Kubis *et al*., 1998; Wicker & Keller, 2007; Wicker *et al*., 2007; Kapitonov & Jurka, 2008; Heslop-Harrison & Schwarzacher, 2011). TEs are divided into two major classes on the basis of their transposition intermediate: Class I TEs or retrotransposons transpose via an RNA intermediate and Class II or DNA transposons transpose directly from DNA with the help of transposase (Finnegan, 1989; Jurka *et al*., 2007; Kaptinov & Jurka, 2008). Retrotransposons or DNA transposons are further classified as autonomous and non-autonomous based on presence or absence of reverse transcriptase (RT) and transposase (TPase) domains respectively. Almost all retrotransposons and DNA transposon superfamilies have autonomous and non-autonomous elements which may differ by as little as a single frame shift or base pair change, or may have major internal deletions (Jurka *et al*., 2007; Wicker *et al*., 2007). The non-autonomous DNA transposons are more abundant as compared to their autonomous counterparts (Rubin *et al*., 2001).

Mutator-like Elements (*MULEs*) are considered to be the most mutagenic plant DNA transposons due to their movement and are widespread among angiosperms. They can capture the genetic sequences of the host and can mobilize the captured fragments to new sites (Diao & Lisch, 2006). *Mutator* transposons are characterized by 9 bp Target Site Duplications (TSDs), 100-500 bp Terminal Inverted Repeats (TIRs) and sometimes have additional direct or indirect repeat sequences in their genomes. TIRs and TSDs are also conserved in these elements and remained constant with continued transposition activity, while internal regions are highly variable with no similarity to each other. The long TIRs of *Mutator* transposons exhibit binding sites and promoters, and hence play a critical role in transposition (Diao and Lisch, 2006; Cifuentes *et al*., 2010). *Pack-MULEs* are the non-autonomous *Mutators* capturing the host genes or gene fragments (Jiang *et al*., 2004). Many *Mutator-like* elements have been investigated, and recently *Pack-MULEs*, which contain fragments of genes, were discovered, distributed among several species of *Gramineae* including wheat, maize, barley, rice, sorghum and bamboo (Lisch *et al*., 2001; Juretic *et al*., 2005).

The genus *Brassica* (family *Brassicaceae*) includes highly valuable crops such as oilseed rape (canola), cabbage, brown mustard, turnip, broccoli, cauliflower, kale, brussels sprouts and kohlrabi, and is a close relative of the model plant *Arabidopsis*. *Brassicas* are a valuable and long-standing food source in both developing and industrialized countries (Monteiro & Lunn, 1999). The evolutionarily-recent polyploidy events led to the formation of three tetraploid species from the diploids *B. rapa* (2n=2x=20), *B. nigra* (2n=2x=16) and *B. oleracea* (2n=2x=18), forming the "Triangle of U" (Nagaharu, 1935), where the three allotetraploid species *B. juncea* (2n=4x=36), *B. napus* (2n=4x=38) and *B. carinata* (2n=4x=34) represent hybrids of each pair of the three diploid species (Fig. 1). The TE content of *B. rapa* was analyzed from 12017 BAC end sequences, which revealed that 14% of the genome is composed of TEs, where retrotransposons are highly abundant (12.3%) in comparison to DNA transposons (1.7%) (Hong *et al*., 2006). Among DNA transposons, *MuDR*, *CACTA* (Alix *et al*., 2008), *hAT*, *Harbinger* and *Mariner* are most common (Kwon *et al*., 2006). In *B. oleracea*, a greater proportion (20%) of TEs was estimated, in contrast to *B. rapa* (Zhang & Wessler, 2005). The present study aimed to identify non-autonomous *Mutator-like* family, which lack any protein coding domains, and hence are difficult to identify, and to investigate the activity of these elements with molecular and bioinformatics methods in diverse *Brassica* germplasms.

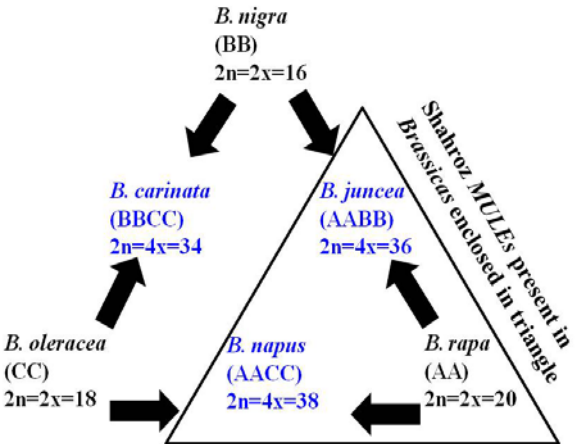

Fig. 1. The relationships and origin of allotetraploids from diploid *Brassica* species of the "Triangle of U" showing distribution of *Shahroz* family of *MULEs*.

Material and Methods

Plant material for *Brassica*: The genomes from 40 *Brassica* accessions (Table 1) were used in the present study. Seeds from 32 *Brassica* accessions were obtained from the Warwick Research Institute (WRI), Warwick, UK, three *B. juncea* (NARC-1, NARC-II and NATCO) and one *B. carinata* (NARC-PK) accession were brought from National Agriculture and Research Center (NARC), Islamabad, Pakistan. Dr. Xian Hong Ge from University of Wuhan, China provided DNA from four synthetic allohexaploids (2n=6x) *Brassica* (Ge *et al*., 2009). The plants were grown in greenhouse at Department of Biology, University of Leicester, UK and DNA was extracted by standard CTAB method (Doyle & Doyle, 1990). Various *Brassica* species with their accession names are listed in Table 1.

Computational analysis for identification and characterization of *MULEs* in *Brassica*: For the *de novo* identification of *Brassica MULEs*, dot plot analysis was performed followed by BLAST searches of the nucleotide sequence database. Homeologous *B. rapa* (AA) and *B. oleracea* (CC) BAC sequences were plotted against each other in JDotter software (Sonnhammer & Durbin, 1995) to find any sequence regions of 100-2000 bp where one BAC had an insertion that was absent from the other (deletion-insertion pairs). The TSDs were investigated in the terminal flanking sequences and TIRs in the insertion sequences. The other homologous copies were analyzed against the NCBI *Brassica* Nucleotide Collection (nr/nt) and Whole-genome Shotgun (WGS) databases using the BLASTN program (Altschul *et al*., 2009). The numbers of strong hits against the reference queries with >70% query coverage and identity were investigated further. The following recommended formula (Tu, 2001) was used to estimate the copy number of *MULEs* in whole *Brassica* genomes: Copy no. = no. of collected *MULEs* in NCBI database x total genome size of *Brassica*/available genome size of *Brassica* in database. For the detailed characterization of *MULEs*, the sequences were investigated for their TSDs, TIRs and internal regions. The sequences were BLAST against the Repbase and Plants Repeat Databases for homology based characterization. The elements not characterized on homology basis were characterized on the basis of their structural hallmarks (TSDs and TIRs, internal non-autonomous regions). The names to the *MULEs* were given systematically according to the recommendations of Capy (2005).

Table 1. List of *Brassica* species with their accession names.

No.	Species	Accession name	No.	Species	Accession name
1.	*B. rapa chinensis*	Pak Choy	21.	*B. juncea*	Tsai Sim
2.	*B. rapa pekinensis*	Chinese Wong Bok	22.	*B. juncea*	W3
3.	*B. rapa chinensis*	San Yue Man	23.	*B. juncea*	Giant Red Mustard
4.	*B. rapa rapa*	Hinona	24.	*B. juncea*	Varuna
5.	*B. rapa rapa*	Vertus	25.	*B. napus*	New
6.	*B. rapa*	Suttons	26.	*B. napus oleifera*	Mar
7.	*B. nigra*	ND	27.	*B. napus biennis*	Last and Best
8.	*B. nigra*	ND	28.	*B. napus napoB.*	Fortune
9.	*B. nigra*	ND	29.	*B. napus*	Drakker
10.	*B. juncea*	NARC-I	30.	*B. napus*	Tapidor
11.	*B. juncea*	NATCO	31.	*B. carinata*	Addis Aceb
12.	*B. juncea*	NARC-II	32.	*B. carinata*	Patu
13.	*B. oleracea gemmifera*	De Rosny	33.	*B. carinata*	Tamu Tex-sel Greens
14.	*B. oleracea*	Kai Lan	34.	*B. carinata*	Mbeya Green
15.	*B. oleracea*	Early Snowball	35.	*B. carinata*	Aworks-67
16.	*B. oleracea italic*	Precoce Di Calabria Tipo Esportazione	36.	*B. carinata*	NARC-PK
17.	*B. oleracea capitata*	Cuor Di Bue Grosso	37.	*B. napus* x *B. nigra*	ND
18.	*B. oleracea*	ND	38.	*B. carinata* x *B. rapa*	ND
19.	*B. juncea*	Kai Choy	39.	*B. napus* x *B. nigra*	ND
20.	*B. juncea*	Megarrhiza	40.	*B. napus* x *B. nigra*	ND

ND: Not Determine. *B. rapa*: AA, *B. nigra*: BB, *B. oleracea*: CC, *B. juncea*: AABB, *B. napus*: AACC, *B. carinata*: BBCC, hexaploid *Brassicas*: AABBCC.

PCR amplification of *Brassica MULEs*: Degenerate primers were designed from the regions flanking the *Brassica* MULE 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 (KAPA, UK), 1.0 mM additional MgCl2, 1 U KAPA Taq DNA polymerase (KAPA, UK), 200-250 mM dNTPs and 0.75 µl (10 pmoles) of each primer. The thermal cycling conditions were adjusted as follows: 3 min denaturation at 94˚C; 35 cycles of 45 sec denaturation at 94˚C, 45 sec annealing at 52-64˚C (depending on primers) and 1 min extension at 72˚C; a final 3 min extension is at 72˚C. PCR products were separated by electrophoresis in 1% agarose gel with TAE buffer according to standard protocols. Gels were stained with addition of 1-2 µl ethidium bromide for the detection of DNA bands under UV illumination.

Sequence analysis, alignment and phylogenetic analysis of *Brassica MULEs*: The GC and AT contents of the *MULEs* were calculated using the online program "GC Calculator" (http://www.genomicsplace.com/gc_calc.html). To detect any protein encoding regions, the sequences were BLAST against the 'Conserved Domain Database (CDD)' available in NCBI. The sequence logos were generated by WebLogo (http://weblogo.berkeley.edu/logo.cgi). The logos representing the insertion preferences for A/T or G/C rich regions were generated by collecting the 30 bp sequences from 5´ and 3´ flanking regions. For the phylogenetic analysis, the TIRs were aligned in Geneious Pro 5.5 software using Neighbor-Joining method with 1000 bootstrap replicates.

Results

Identification of non-autonomous *Mutator-like* elements (*MULEs*) in *Brassica*: The first *Mutator-like* element (*BrN-MULE1-1*) in *Brassica* was identified by the comparison of two homoeologous BACs: *B. rapa* (AC189298.1) against *B. oleracea* (EU642504.1) (Fig. 2a). The element was 2781 bp long, flanked by 9 bp TSDs, exhibit 76 bp imperfect TIRs with no internal coding region (non-autonomous). After the identification of *BrN-MULE1-1*, the comparative analysis of *B. rapa* accession (AC155342.2) against its homologue *B. rapa* (AC146875.2) (Fig. 2b) led to the discovery of a 2920 bp element (*BrN-MULE1-2*). The elements showed high AT contents in their internal regions, which are not observed in other *MULEs* investigated in rice, maize and *Arabidopsis* plants. It showed typical features of *MULEs* by exhibiting 9 bp TSDs and 76 bp TIRs. The initially identified elements were BLAST against *Brassica* nucleotide collection and whole-genome shotgun sequences databases in GenBank and complete sequences were retrieved. No homology was detected with other known TEs by blasting against Repbase, and Plant Repeat databases of transposable elements. Based on novelty of these elements, we report this new *MULE* family and named it *Shahroz* family of *MULEs*.

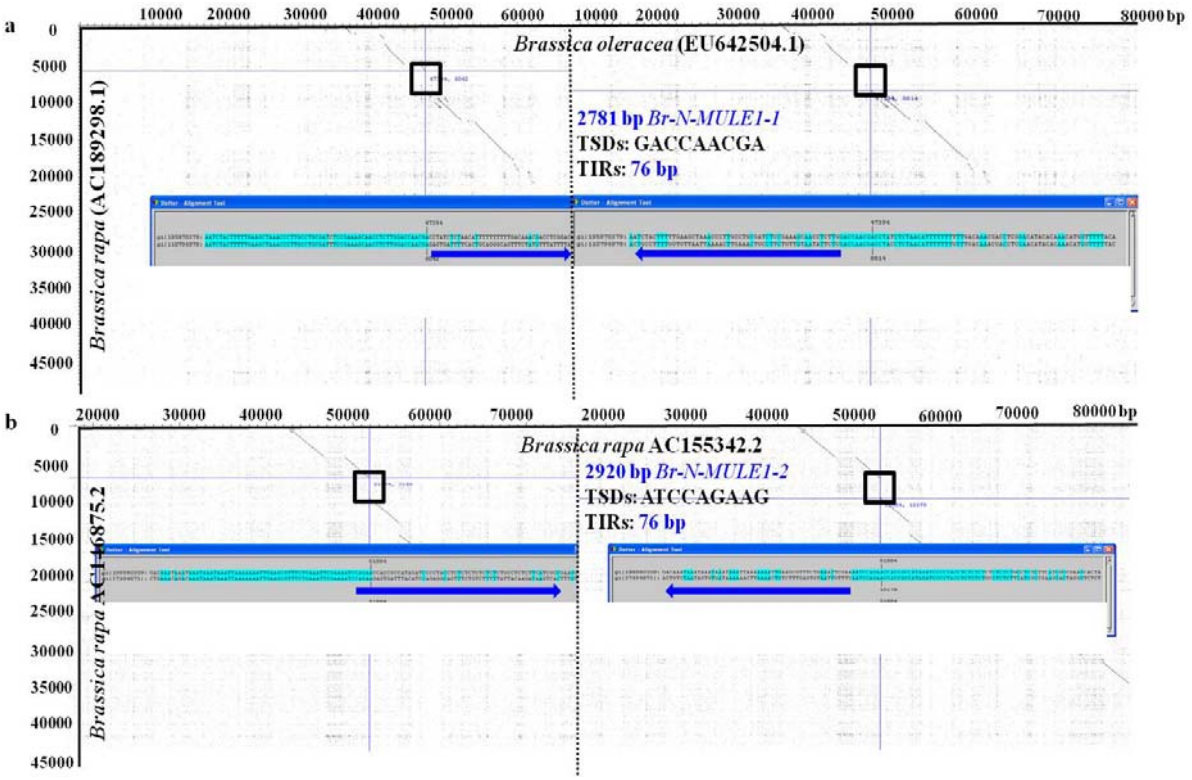

Fig. 2. Dot plot comparison of homoeologous BAC clones of *Brassica* identified a) *BrN-MULE1-1* b) *BrN-MULE1-2* insertion sites in *Brassica* BAC accessions. The size, TSDs and TIRs are also indicated. The opposing arrows are indicating the TIRs on the sequence alignment insets.

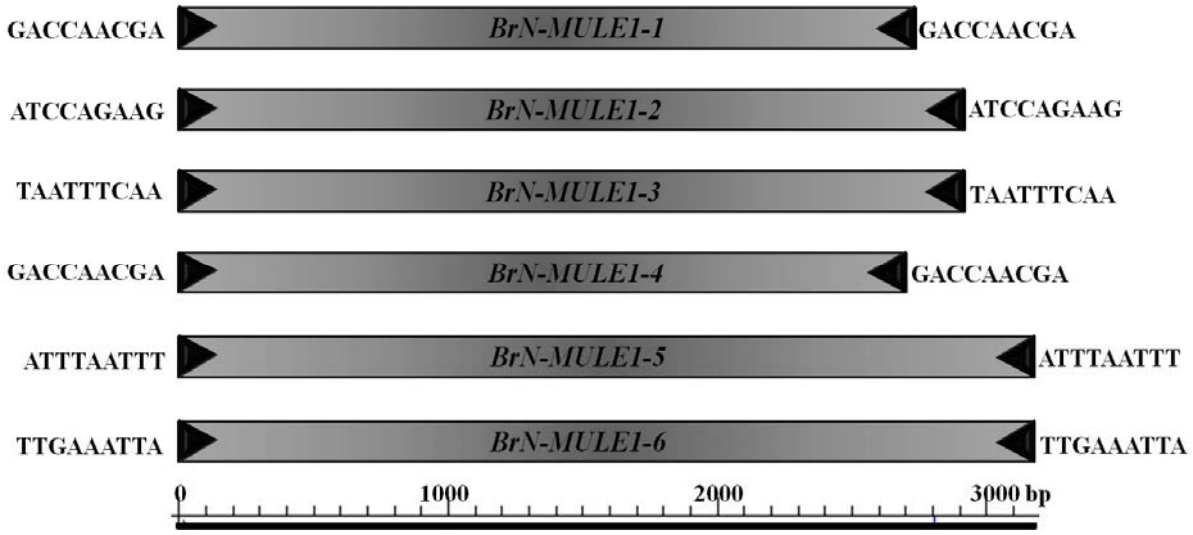

Fig. 3. Schematic representations of *Brassica* non-autonomous *Mutator-like* elements from *Shahroz* family. Black triangles represent 76 bp TIRs. The 9 bp TSD sequences are shown at both termini. The scale represents size in bp.

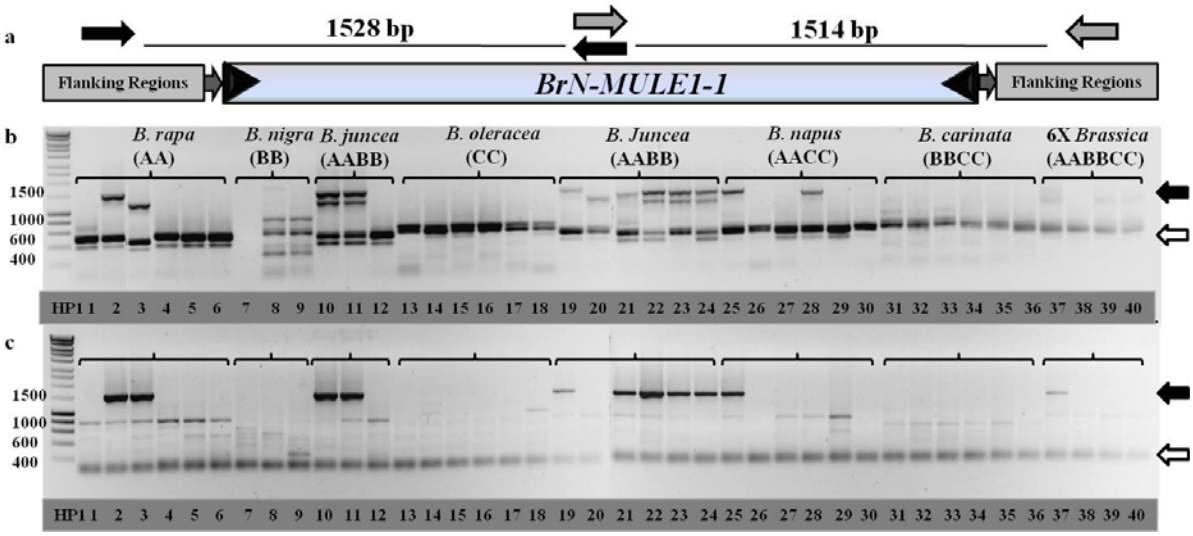

Fig. 4. Insertion polymorphisms of *Mutator* elements in *Brassica*. a) Schematic representation of *BrN-MULE1-1* showing the position of forward and reverse primers (arrows); b & c) The amplification of *BrN-MULE1-1* insertion sites in various *Brassica* accessions (Table 1)*:* Long bands (1528/1514 bp; filled arrows) show the amplified element and short bands indicated by empty arrows amplify the empty sites only. The figure shows 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. Left lane (HP1) is 200 bp marker ladder (Hyperladder I) with band sizes indicated.

Molecular characterization of *Brassica MULEs*: The structural features of *Shahroz* family of *MULEs* identified from *Brassica* revealed that the elements range in sizes from 2734-3160 bp. The identified elements shared 52-90% sequence identity at the nucleotide level. The full length non-autonomous *MULEs* copies were confirmed by presence of TIRs at both ends followed by multiple sequence alignments. The first element *BrN-MULE1-1* was 2781 bp in size including 9 bp TSDs (5´-GACCAACGA-3´) and displayed 76 bp imperfect TIRs (Fig. 3). The internal region of the element is AT rich with only 37% GC content. *BrN-MULE1-2* was 2920 bp large in size, including 9 bp TSDs (ATCCAGAAG) and 76 bp imperfect TIRs. BLASTN searches resulted in the collection of 4 other homologous of similar sizes (~2.7-3.1 kb). *BrN-MULE1-3* is homologous (99%) to *BrN-MULE1-2* except exhibiting distinct TSDs and was identified from *B. rapa* accession (AC189583.2) at nucleotide position 42105-45024 bp in BAC sequence. It is flanked by 9 bp TSDs (TAATTTCAA) and 76 bp TIRs, which are highly conserved among the two sequences (Fig. 3; Table 2). *BrN-MULE1-4* is 2734 bp large element including the 9 bp TSDs and exhibiting 76 bp imperfect TIRs. The sequence was retrieved from Whole-genome Shotgun sequences database of GenBank residing in *B. rapa* accession AENI01006341.1. The hits showed the localization of *BrN-*

MULE1-4 on chromosome number 7 and 8 of *B. rapa*. Two elements designated as *BrN-MULE1-5* and *BrN-MULE1-6* of 3160 bp sizes were identified from 'AENI01003197.1' and 'AENI01009183.1' accessions indicating their presence on *B. rapa* chromosome 4 and 10 respectively (Table 2). Both elements have shown the canonical non-autonomous *Mutator* structures by displaying 9 bp TSDs and long TIRs (76 bp) (Fig. 3). Their TSDs were variable but TIRs showed the conserved regions indicating the members of the same *Mutator* family.

PCR analysis revealed A-genome specificity of *Shahroz MULEs* family: The genome specificity of some transposons showed high proliferation in one organism but less abundance in their relative species. Our data confirmed the distribution of *Shahroz* family of *MULEs* in A-genome *Brassica* by dot plot, BLAST searches, where output found six copies from *B. rapa* only and molecular analysis yielded the PCR products in A-genomes and its allopolyploids (AABB, AACC, AABBCC). To amplify a 3.2 kb product, the element was split into two fragments. Two primer pairs were designed to amplify 1528 and 1514 bp PCR products from 5´ and 3´ ends respectively (Fig. 4a). The primer pair MULE5´F GAACATGGTCACCTTCACTG + MULE5´R CATGGTTAGAAACCGTGTGG and MULE3´F CCACACGGTTTCTAACCATG + MULE3´R ACGGGGAAATGAAACTGTAG successfully amplified the 1528 and 1514 bp products respectively from *B. rapa* and its allotetraploids, but no amplification in C-genome suggested its absence in *B. oleracea* and its allotetraploids (BBCC).

A total of 40 *Brassica* accessions were tested to amplify the *Shahroz* family of *MULEs* (Figs. 4b, 4c). Of the 40 accessions, 6 were each from A and C-genomes, 3 were from B-genome, 9 were AABB, 6 were each AACC and BBCC and 4 were from AABBCC genomes (see Table 1). The 5´-1528 and 3´-1514 bp fragments were amplified from two *B. rapa* (Chinese Wong Bok, San Yue Man) accessions among A-genomes. All the six *B. rapa* accessions generates 2 additional bands of ~450 and 380 bp. *B. nigra* (B-genome) accessions showed no amplification of the insertion. Nine *B. juncea* cultivars yielded amplification of both 5´-1528 bp and 1514-3´ products. Among six *B. napus*, 2 cultivars (New, Fortune) amplified the bands. From the four synthetic allohexaploids (AABBCC), 3 yielded weak bands of expected sizes. The insertion polymorphisms of *Shahroz* family suggest its distribution in A-genome diploids and their allotetraploids but their absence in B and C-genome.

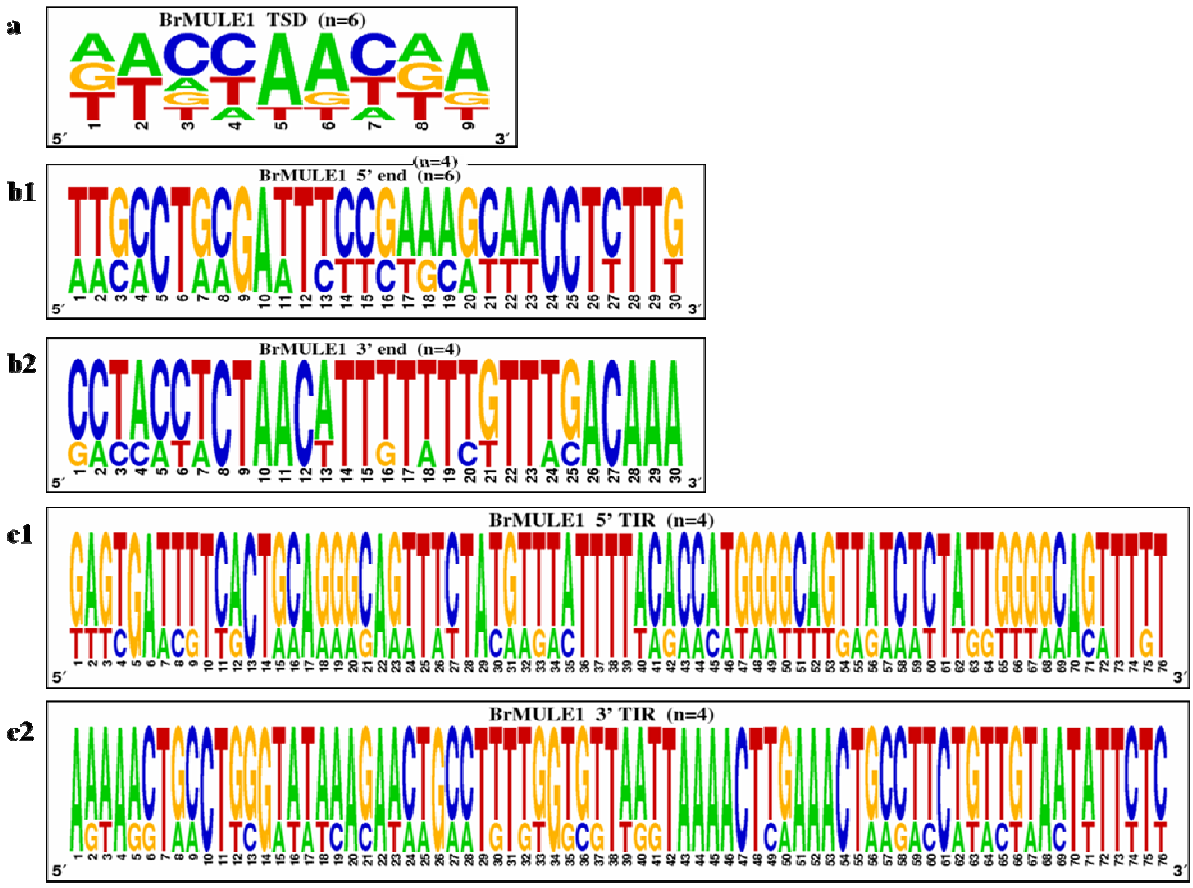

Fig. 5. WebLogos showing the sequence conservation and information content of *Shahroz* family of *Mutator* DNA transposons indicating a) weakly conserved TSDs; b1 & b2) Insertion preference of *Brassica MULEs* in AT rich regions; c1 & c2) strong conservation of half the bases (31 out of 76) in TIRs, with the TA conserved motifs dispersed in the 5' and 3' TIRs. The height of nucleotides indicates the information content of consensus nucleotides at each position.

Table 2. Non-autonomous *Mutator* transposons with accessions numbers, sizes, TSDs, TIRs, positions in the BAC sequences.

Name	Family	BAC Accession	Species	Size	TSD	TSD sequence	TIR	Position
BrN-MULE1-1	*Shahroz*	AC189298.1	*B. rapa*	2781	9	GACCAACGA	76	6034-8806
BrN-MULE1-2	*Shahroz*	AC146875.2	*B. rapa*	2920	9	ATCCAGAAG	76	7268-10179
BrN-MULE1-3	*Shahroz*	AC189583.2	*B. rapa*	2920	9	TAATTTCAA	76	42105-45024
BrN-MULE1-4	*Shahroz*	AENI01006341.1	*B. rapa*	2734	9	GACCAACGA	76	227056-229789
BrN-MULE1-5	*Shahroz*	AENI01003197.1	*B. rapa*	3160	9	ATTTAATTT	76	23051-26210
BrN-MULE1-6	*Shahroz*	AENI01009183.1	*B. rapa*	3160	9	TTGAAATTA	76	751-3911

Insertion preference of *Shahroz* family of *MULEs*: The insertion preferences of *Brassica MULEs* were investigated by analyzing the flanking regions around the insertions. The TSDs of the elements were variable among all *Brassica MULEs* but mostly AT rich (Fig. 5a). We found that *MULE* insertions preferred the AT rich regions in their flanking regions (Figs. 5b1, 5b2) but the percentage of AT was not very high as observed in other TEs and *MITEs*. The 30 bp flanking the site of insertion showed AT content of 55% and 65% at 5´-and 3´ end of insertion respectively. Nearly half of the basis on 3´ flanking region of the insertion were G/C, while GC/AT ratio was very low at 5´ flanking region. The 3´ flanking region started with TT dinucleotide, while 5´ started with CC dinucleotide after the TSD. The starting parts of both 5´ and 3´ flanking regions were GC rich in contrast to the following part (AT rich). A stretch of 10 thymine basis (starting from 14-24) was observed in 3´ flanking region (Fig. 5b2) suggesting the high AT content in 3´ flanking region.

***Shahroz;* a defective and non-autonomous family of *MULEs*:** The comparative analysis of six *Brassica MULEs* showed varied TSDs and homology ranged from 60-97% in their TIRs and 52-55% in their internal regions. There were many nucleotide substitutions as well as conserved and varied regions within TIRs of these elements (Figs. 5c1, 5c2). The 3 bp termini (GAG) at 5´ and 4 bp (TTTT) at 3´ end at position 72-76 bp were highly conserved with some dispersed T and G rich motifs (Fig. 5c1). The BLAST results against known protein database yielded no hits against any transposase or protein coding regions indicating that *Brassica MULEs* are non-autonomous or defective elements of active *Mutator* elements.

Phylogenetic analysis of *Brassica* and other plant *Mutator* elements: The phylogenetic relationship of *Shahroz* family of *MULEs* and other plant *Mutator* elements was studied by Neighbor-Joining method with 1000 bootstraps. The tree was rooted with *Drosophila elegans Mutator* element *MuDR-1*. The 75 bp 5´-TIRs from 20 elements other than *BrN-MULE1* were collected from Repbase database and aligned in Geneious alignment. No species specific clades of *Mutator* elements were observed but we found the clustering of dicot and monocot *Mutator* elements (Fig. 6). Six clades were evident with elements ranging in number from 1 to 6. The *Brassica BrN-MULE1* element shared no family with any known element but out branched showing a separate evolutionary line. *Medicago truncatula* and *Solanum tuberosum Mutator*s shared a clade named RAS1. Another clade designated as CLOUD further splits into three sub-clades composed of *Oryza*, *Triticum* and *Hordeum Mutator* elements with a *Glycine max* MuDr element. *Arabidopsis thaliana VANDLA18* and *ARNOLDY1* elements come close to each other constituting the same clade designated *ARNOLDY1*. Five elements including two from *Arabidopsis* (VANDAL1, ATMU1), one from *Zea mays* (Jittery), one from *Medicago trancatula* (MuDi) and one from *Cucumus melo* (CUMULE) shared the clade VANDAL1, where *Jittery* element out grouped from *Arabidopsis Mutator*s. The MDM1 clade was composed of a *Sorghum bicolour* (*MuDR10_SB*), two *Oryza sativa* (*MUDROS1, MDM1*) and one *Malus domestica* element. The phylogenetic analysis showed no homology of *Brassica MULEs* with other known *Mutator* elements suggesting that the *BrMULEs* are *Brassica* genome specific.

Discussion

Genome sequencing is enabling the characterization of the nature and abundance of the full range of transposon families. The information allows the study of their activity and proliferation in over shorter and longer timescales in germplasms and species collections (Fig. 1). Mobile insertions can be inserted to or near genes and can alter their function as observed by small *SINEs* and *MITE-like* elements (Deragon *et al*., 2008; Feschotte, 2008) in plants, *MULEs* are considered to be one of the elements involved in capturing the genes or gene fragments and hence related to the evolution of genomes (Jiang *et al*., 2011).

In current study, a non-autonomous *Mutator* family, named *Shahroz* was characterized by its evolutionary activity in A- compared to C-genome (Fig. 2a), and between accessions of the A-genome *B. rapa* (Fig. 2b). DNA transposons are proliferating in *Brassica* genomes such as *CACTA* (Alix *et al*., 2008), *Harbingers, hAT* and *MITEs* (Nouroz, 2012; Nouroz *et al*., 2015). *Mutator-like MITEs* with 9-10 bp TSDs and long TIRs were characterized from *Brassica* genomes with sizes <1.6 kb, high AT rich sequences and short internal regions flanked by long TIRs (Zhang & Wessler, 2005; Hong *et al*., 2006). Our *MULEs* can be easily distinguished from the *Mutator-like MITEs* by having large sizes (Fig. 3; Table 2), low copy numbers, short TIRs (76 bp) in contrast to TIRs of *Mutator*-derived *MITEs* (>200 bp), and large internal regions. Insertion preference of *Mutator*-derived *MITEs* was to AT rich genomic regions contrasted to the non-autonomous *Brassica MULEs*, which showed less preference for AT rich regions. Low copy numbers of *MULEs* were noted in other plants like *Zea mays* with <5 copies in each family (Li *et al*., 2009).

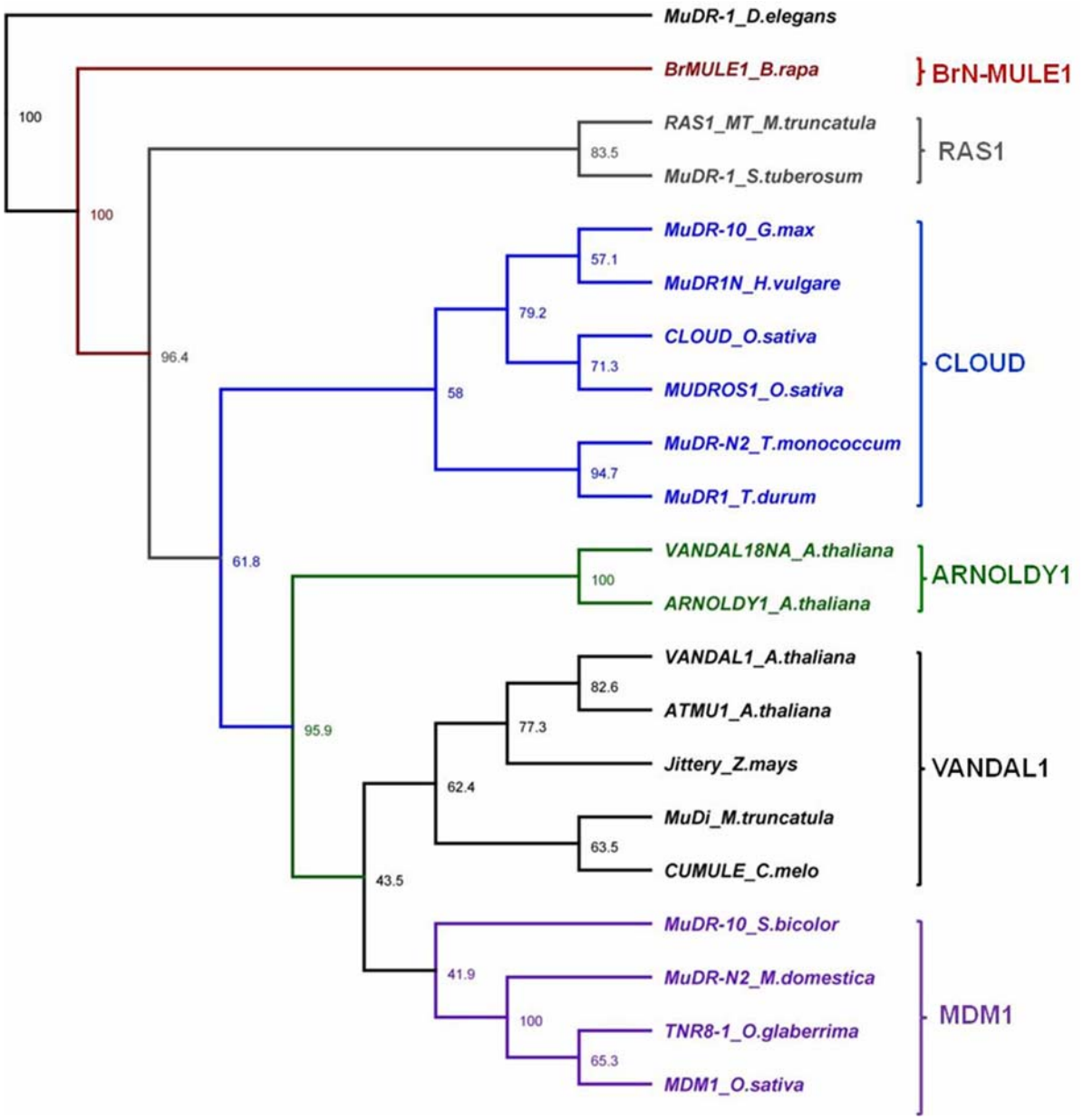

Fig. 6. A Neighbor-Joining tree for phylogenetic analysis of *Brassica* and other plant *Mutator* TEs. The tree uses 1000 bootstrap replicates (consensus indicated as %) and was rooted with *Drosophila elegans MuDr* element. The known *Mutator* sequences were collected from Repbase database of Eukaryotic transposable elements. Six clades are represented by 21 elements.

The insertion polymorphisms of *Mutator* elements across 40 *Brassica* accessions showed its amplification in two *B. rapa* accessions and around ten allotetraploids (Figs. 4b, 4c), all with the A-genome, suggesting that it is contributed from A-genome. This amplification among two *B.rapa* accessions and their absence from other accessions showed intra-specific variation. Such molecular markers are highly informative to study the biodiversity and evolution of plant genomes. Molecular markers such as Retrotransposon Based Insertion Polymorphism (RBIP) (Flavell *et al*., 1998), Inter-Retrotransposon Polymorphism (IRAP), Retrotransposon-Microsatellite Polymorphism (REMAP) (Kalendar & Schulman, 2006) and *MITEs* insertion polymorphism markers (Yaakov *et al*., 2012) are extensively in use to study the biodiversity and evolution in plant genomes. As evolutionarily active elements, *MULE* sequences and primers used here are also useful for measurement of biodiversity and phylogeny within *Brassica*.

The phylogenetic analysis of plant *MULEs* revealed that *BrN-MULEs* are a well-supported sister group to *Mutator* elements from other plants (Fig. 6). Other clades of *Mutator* elements showed some bootstrap support, and included both monocot and dicot elements. Thus the *Shahroz* family of *MULEs* in *Brassica* is distinct from other *MULEs*; so it will be interesting to find whether the diversity of all *MULEs* in plants, including low-copy number sequences, is higher than previously suggested.

The internal regions of the *Shahroz MULEs* were variable with degrading copies. TIRs of *MULEs* are responsible for the transposition activity of the elements, but the mutated TIRs with small insertions or substitutions are probably inactive elements (Jiang *et al*., 2011). The detailed analysis of *Mutator-like* elements in maize, tomato and *Arabidopsis* genomes showed evidence of non-autonomous copies of *MULEs* (Cifuentes *et al*., 2010). Many non-autonomous TEs still retain the sequences required for the recognition of transposase and hence are mobilized *in trans* by autonomous elements residing nearby (Hartl *et al*., 1992). Almost all TE superfamilies have acquired non-autonomous elements as first identified element activator (McClintock, 1950) also have its non-autonomous partner named Dissociation (Ds). The low copy numbers of Pack-*MULEs* in plant genomes by Jiang *et al*. (2011) is supported by our analysis, where we counted low copies of *MULEs* in *Brassicaceae*. The relatively short, degrading nature of TIRs, low copy numbers and lack of non-coding regions indicate the ancient nature of *Shahroz* family. Such defective and internally deleted DNA transposons are studied in several organisms such as non-autonomous *hAT*, *CACTA*, and *Mutator* investigated in plant, animal and fungal genomes (Benjak *et al*., 2009; Rubin *et al*., 2001). The degeneracy and non-autonomous nature of the *Shahroz MULE* family in *Brassica* suggests that it is an ancient genomic component now passing into the evolutionary stages of degradation.

Conclusion

Transposable elements constitute the major components of plant genomes. In the present study, the analysis helped us in the identification and characterization of less active non-autonomous *Mutator-like* family named *Shahroz* and its amplification polymorphisms in diverse *Brassica* germplasms. The results suggested the separate line of evolution for *Brassica Shahroz* family of *MULEs* as compared to other plant *MULEs*. The markers were informative to study the *MULEs* insertion polymorphisms of *Brassica* genomes and developing of similar markers from related transposons families will be more helpful in intra-specific biodiversity and evolution.

Acknowledgements

We are grateful to Hazara University Mansehra, Pakistan 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. Xianhong Ge from University of Wuhan, China for supplying seeds or DNA from the *Brassica* accessions studied.

References

Alix, K., J. Joets, C. Ryder, J. Moore, G. Barker, J. Bailey, G. King and J.S. Heslop-Harrison. 2008. The CACTA transposon Bot1 played a major role in *Brassica* genome divergence and gene proliferation. *Plant J.*, 56: 1030-1044.

Altschul, S.F., E.M. Gertz, R. Agarwala, A.A. Schaffer and Y.K. Yu. 2009. PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res.*, 37: 815-824.

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

Capy, P. 2005. Classification and nomenclature of retrotransposable elements. *Cytogenet. Genome Res.*, 110: 457-461.

Cifuentes, M., F. Eber, M-O. Lucas, M. Lode, A-M. Chevre and E. Jenczewski. 2010. Repeated polyploidy drove different levels of crossover suppression between homoeologous chromosomes in *Brassica napus* allohaploids. *Plant Cell*, 22(7): 2265-2276.

Deragon, J.M., J.M. Casacuberta and O. Panaud. 2008. Plant transposable elements. *Genome Dyn.*, 4: 69-82.

Diao, X.M. and D. Lisch. 2006. *Mutator* transposon in maize and *MULEs* in the plant genome. *Acta Genet. Sinica*, 33(6):477-487.

Doyle, J.J. and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. *Focus*, 12: 13-15.

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

Finnegan, D.J. 1989. Eukaryotic transposable elements and genome evolution. *Trends Genet.*, 5(4): 103-107.

Flavell, A.J., M.R. Knox, S.R. Pearce and T.H. Ellis. 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. *Plant J*., 16: 643-650.

Ge, X.H., J. Wang and Z.Y. Li. 2009. Different genome-specific chromosome stabilities in synthetic *Brassica* allohexaploids revealed by wide crosses with *Orychophragmus*. *Ann. Bot.*, 104: 19-31.

Hartl, D.L., E.R. Lozovskaya and J.G. Lawrence. 1992. Nonautonomous transposable elements in prokaryotes and eukaryotes. *Genetica*, 86(1-3): 47-53.

Heslop-Harrison, J.S. and T. Schwarzacher. 2011. Organisation of the plant genome in chromosomes. *The Plant J.*, 66(1): 18-33.

Hong, C.P., P. Plaha, D.H Koo, T.J. Yang, S.R. Choi, Y.K. Lee, T. Uhm, J.W. Bang, D Edwards, I. Bancroft, B.S. Park, J. Lee and Y.P. Lim. 2006. A survey of the *Brassica rapa* genome by BAC-end sequence analysis and comparison with *Arabidopsis thaliana*. *Mol. Cells*, 22 (3): 300-307.

Jiang, N., Z. Bao, X. Zhang, S.R. Eddy and S.R. Wessler. 2004. Pack-mule transposable elements mediate gene evolution in plants. *Nature*, 431(7008): 569-573.

Jiang, N., A.A. Ferguson, R.K. Slotkin and D. Lisch. 2011. Pack-*Mutator-like* transposable elements (Pack-*MULEs*) induce directional modification of genes through biased insertion and DNA acquisition. *Proc. Natl. Acad. Sci. USA*, 108(4): 1537-1542.

Juretic, N., D.R. Hoen, M.L. Huynh, P.M. Harrison and T.E. Bureau. 2005. The evolutionary fate of mule-mediated duplications of host gene fragments in rice. *Gen. Res.*, 15(9): 1292-1297.

Jurka, J., V.V. Kapitonov, O. Kohany and M.V. Jurka. 2007. Repetitive sequences in complex genomes: Structure and evolution. *Annu. Rev. Genomics Hum.Genet.*, 8: 241-259.

Kalendar, R. and A.H. Schulman. 2006. IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat. Protoc.*, 1(5): 2478-2484.

Kapitonov, V.V. and J. Jurka. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet*., 9: 411-412.

Kubis, S.E., J.S. Heslop-Harrison, C. Desel and T. Schmidt. 1998. The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. *Plant Mol. Biol.*, 36: 821-831.

Kwon, S.J., S.W. Hong, J.H. Son, J.K. Lee, Y.S. Cha, M.Y. Eun and N.S. Kim. 2006. CACTA and MITE transposon distributions on a genetic map of rice using F15 RILs derived from Milyang 23 and Gihobyeo hybrids. *Mol. Cells*, 21(3):360-366.

Li, Q., L. Li, J. Dai, J. Li and J. Yan. 2009. Identification and characterization of CACTA transposable elements capturing gene fragments in maize. *Chinese Sci. Bull.*, 54(4): 642-651.

Lisch, D.R., M. Freeling, R.J. Langham, M.Y. Choy. 2001. *Mutator* transposase is widespread in the grasses. *Plant Physiol.*, 125(3): 1293-1303.

McClintock, B. 1950. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA*, 36(6): 344-355.

Mills, R.E., E.A. Bennett, R.C. Iskow, C.T. Luttig, C. Tsui, W.S. Pittard and S.E. Devine. 2006. Recently mobilized transposons in the human and chimpanzee genomes. *Am. J. Hum. Genet.*, 78(4): 671-679.

Monteiro, A. and T. Lunn. 1999. Trends and perspectives of perspectives of vegetable *Brassica* breeding world-wide. *Acta Hort.*, 495: 273-280.

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

Nouroz, F. 2012. The structures and abundance of transposable elements contributing to genome diversity in the diploid and polyploid *Brassica* and *Musa* crops. PhD Thesis, University of Leicester, UK.

Nouroz F., S. Noreen and J.S. Heslop-Harrison. 2015. Evolutionary genomics of miniature inverted-repeat transposable elements (MITEs) in *Brassica*. *Mol. Genet. Genomics*, 290: 1076-1089.

Rubin, E., G. Lithwick and A.A. Levy. 2001. Structure and evolution of the hAT transposon superfamily. *Genetics*, 158: 949-957.

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

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.

Wicker, T. and B. Keller. 2007. Genome-wide comparative analysis of copia retrotransposons in *Triticeae*, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. *Genome Res.*, 17: 1072-1081.

Wicker, T., F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A.H. Schulman. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.*, 8(12): 973-982.

Yaakov, B., E. Ceylan, K. Domb and K. Kashkush. 2012. Marker utility of miniature inverted-repeat transposable elements for wheat biodiversity and evolution. *Theo. Appl. Genet.*, 124(7): 1365-1373.

Zhang, X. and S.R. Wessler. 2005. BoS: A large and diverse family of short interspersed elements (SINEs) in *Brassica oleracea*. *J. Mol. Evol.*, 60(5): 677-687.

(Received for publication 28 June, 2014)