Pak. J. Bot., 49(2): 789-798, 2017.

IDENTIFICATION AND EVOLUTIONARY DYNAMICS OF *CACTA* DNA TRANSPOSONS IN *BRASSICA*

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

[1]*Department of Botany, Hazara University Mansehra, Pakistan*
[2]*Molecular Genetics Laboratory, Department of Genetics, University of Leicester, UK*
[3]*Molecular Cytogenetics Laboratory, Department of Biology, University of Leicester, UK*
[*]*Corresponding author's e-mail: faisalnouroz@gmail.com. +92 997414168*

Abstract

Transposable elements are the major drivers of genome evolution and plasticity. Due to their transposition mode, they are classified into two major classes as Retrotransposons and DNA transposons. The *En/Spm* or *CACTA* elements are diverse group of DNA transposons proliferating in plant genomes. Various bioinformatics and molecular approaches were used for identification and distribution of *CACTA* transposons in *Brassica* genome. A combination of dot plot analysis and BLASTN searches yielded 35 autonomous and 7 non-autonomous *CACTA* elements in *Brassica*. The elements ranged in sizes from 1.2 kb non-autonomous elements to 11kb autonomous elements, terminated by 3 bp Target Site Duplication (TSD) and ~15 bp conserved Terminal Inverted Repeat (TIR) motifs (5′-*CACTA*CAAGAAAACA-3′), with heterogeneous internal regions. The transposase (TNP) was identified from autonomous *CACTA* elements, while other protein domains from *Brassica* and other plants *CACTA* revealed similar organizations with minor differences. Both transposases (TNPD, TNPA) are present in most *CACTA*, while a few *CACTA* harboured an additional ATHILA ORF1-like domain. The PCR analysis amplified the *CACTA* transposases from 40 *Brassica* accessions (A, B, and C-genome) suggesting their distribution among various *Brassica* crops. A detailed characterization and evolutionary analysis of the identified *CACTA* elements allowed some to be placed in genome-specific groups, while most of them (*Brassica-Arabidopsis* elements) have followed the same evolutionary line. The distribution of *CACTA* in *Brassica* concluded that 3 bp TSDs generating *CACTA* transposons contributed significantly to genome size and evolution of *Brassica* genome.

Key words: *Brassica*, DNA transposons, *CACTA*, Autonomous, Transposase, Phylogenetic analysis.

Introduction

Brassica a highly diverse genus of family Brassicaceae is economically very important due to valuable crops such as Chinese cabbage, broccoli, cauliflower, brussels sprouts, collards, turnip, brown mustard and oilseed rape (canola) used for vegetables, oils, forage and as ornamentals. *Brassica rapa* (AA), *B. nigra* (BB) and *B. oleracea* (CC) are three diploid species, which yielded 3 allotetraploid species as *B. juncea* (AABB), *B. napus* (AACC) and *B. carinata* (BBCC) by their hybridization and polyploidization (Monteiro & Lunn, 1999; Christopher *et al*., 2005). *B. oleracea* is comprised of several important crops, although 10 other non-domesticated genotypes have also been described (Ostergaard & King, 2008). *Brassica* genome has shown high similarity in genome of its allied genus *Arabidopsis*. The divergence of *Brassica-Arabidopsis* lineage is estimated between 14.5-24 million years ago (Mya) (Yang *et al.,* 1999; Koch *et al.,* 2000). As identified in other plants, the genome of *Brassica* also harbour transposable elements (TEs) such as LTR retrotransposons (Nouroz *et al*., 2015a), DNA transposons like *Mutator* (Nouroz *et al*., 2015b), *hATs* (Nouroz *et al*., 2015c) and *Harbingers* (Zhang & Wessler, 2004; Nouroz *et al*., 2016).

Successful crops have shown genetic diversity and variations in their genomes (Sidra *et al*., 2014). Like other factors, transposable elements are major drivers of genome evolution and diversity. Among them, Class I or DNA-mediated transposons adopted "cut and paste" mechanism of transposition and are characterized by possessing target site duplication (TSD), terminal inverted repeats (TIRs), and a DDD or DDE motif specific transposase required for their transposition. Based on structural diversity of transposases, DNA transposons are clustered in to several families, of which 6 (*CACTA, hAT, Harbingers, Helitron, Mutator* and *Mariner*) are common in plants (Wicker *et al*., 2007; Kapitonov & Jurka, 2008). *En/Spm* or *CACTA* elements are autonomous DNA transposons with an active transposase, while their non-autonomous partners I/dSpm (Inhibitor/dSpm) lack the transposase. The non-autonomous inhibitor and the defective *Spm* (*dSpm*) are the deletion derivatives of the autonomous elements. *CACTA* elements are named on the basis of their conserved DNA sequence signature in termini of their TIRs (Pereira *et al*., 1986) and are flanked by 3 bp TSDs, 10-28 bp TIRs and DDD/E type transposase (Wicker *et al*., 2003; Tian, 2006).

Many families of the *CACTA* superfamily have been described from the grass family as *Baldwin, Casper, Enac, Isaac, Jorge, Mandrake* and *TAT-1*. The internal sequences of the elements are highly divergent; although 20-30 bp TIRs including the *CACTA* motif are similar. They are not easily identified by computer aided database searches (Wang *et al*., 2003; Wicker *et al*., 2003). Generally, the autonomous *CACTA* contain the *CACTA* transposase protein (TNPD) responsible for transposition and integration, while another transposase protein (TNPA) is a factor performing multiple functions (Trentmann *et al*., 1993; Gierl, 1996).

CACTA elements constitute a diverse group of DNA transposons identified from various plants and include *Caspar* from *Triticum* (Sergeeva *et al*., 2010), *Tam1* and *TamRS1* from snapdragon (*Antirrhinum majus*) (Roccaro *et al*., 2007), *En/Spm* and *dSpm* from maize (Gierl, 1996), *CAC1* from *Arabidopsis thaliana* (Miura *et al*., 2001) and *Caspar* from *Triticeae* (Wicker *et al*., 2003). The numbers and distribution patterns of *CACTA* DNA transposons vary in various genomes. Of the DNA transposons investigated in common bean (*Phaseolus vulgaris*),

maximum copies of *CACTA* (348) were identified followed by *MULEs* (45), *Helitrons* (39) and *hATs* (23), which indicated that some genomes have more plasticity of *CACTA* elements to proliferate (Gao *et al*., 2014). The *CACTA* elements are used as molecular markers in many crops as in maize, the markers were developed from TIRs of *Issac-CACTA* transposons to distinguish the maize inbreed lines (Lee *et al.*, 2005).

We aimed here to identify the *CACTA* transposons in *Brassica* genome and to analyse their structures, the evolutionary diversity, mobility and consequences for *Brassica* genome organization. *Brassica CACTA* elements were compared with *CACTA* from other crops to observe the structural diversity and evolutionary relationships.

Material and Methods

Plant material for *Brassica*: Of the 40 *Brassica* accessions/genotypes (Table 1) used in present study, seeds from 32 *Brassica* accessions were brought from Warwick Research Institute (WRI), UK, 4 from National Agriculture and Research Centre (NARC), Islamabad, Pakistan and DNA for 4 synthetic allohexaploids (2n=6x) *Brassica* (Ge *et al.*, 2009) were provided by Xian Hong Ge (University of Huazhong Agricultural University, Wuhan, China). The standard CTAB method (Doyle & Doyle, 1990) was used for DNA extraction from young fresh leaves grown in green house of Department of Biology, University of Leicester, UK.

Computational analysis for characterization of *Brassica CACTA*: Dot plot analyses were performed for *de novo* identification of *Brassica CACTA* elements. Homeologous *Brassica rapa* (AA) and *Brassica oleracea* (CC) bacterial artificial chromosome (BAC) sequences were plotted against each other in JDotter software (Sonnhammer & Durbin, 1995) to find any deletion-insertion pairs where one BAC had a sequence fragment that was absent from the other. The TSDs were investigated manually in the terminal flanking sequences and TIRs in the insertion sequences. The other homologous copies were collected against the NCBI *Brassica* Nucleotide collection (http://www.ncbi.nlm.nih.gov) using BLASTN program. The elements were characterized on the basis of their structural hallmarks (TSDs, TIRs, transposase and associated domains) into their respective superfamily and families. To detect the protein encoding domains, the sequences were screened against the Conserved Domain Database (CDD) available in NCBI.

Naming the *CACTA* transposons: The names to the *CACTA* were given as ***BoCACTA1-1***, where '***B***' stands for genus *Brassica*, second letter '***o'*** represents *oleracea*, 5 capital letters '***CACTA***' represent the transposons superfamily, the first number indicate the family and number followed by hyphen represents number of the respective member of family. For non-autonomous elements letter '***N***' is used before superfamily to indicate a non-autonomous element.

PCR amplification of *Brassica CACTA* transposase: To amplify *CACTA* transposase, degenerate primers pair BoCACTAF (5′-CCTCAGGTGGACCATCAAAC-3′) and BoCACTAR (3′-GACGAAAAGGTTGCAGAGGT-5′) was designed from the conserved DDD/E triad motifs of transposase (TNPD) by using Primer3 (http://frodo.wi.mit.edu/primer3/). For amplification of ATHILA domain, the primers BoATHILAF (5′-ACATTGAAGGGCTGTTCCAG-3′) and BoATHILAR (3′-AGCTTGTACTGGCTGGAGTC-5′) were designed. PCR amplifications were done by using 50 ng *Brassica* DNA in 15 µl reaction mixture with 2 µl PCR buffer (KAPPA, UK), 1.0 mM MgCl2, 200-250 mM dNTPs, 0.75 µl of each primer and 1U KAPPA Taq polymerase (KAPPA, UK). The thermal cycling conditions were 3 min denaturation at 94˚C; 35 cycles of 45 sec denaturation at 94˚C, 45 sec annealing at 58-60˚C, 1 min extension at 72˚C and final 3 min extension at 72˚C. PCR products were separated by electrophoresis in 1% agarose gel with TAE buffer according to the standard protocols. Gels were stained with addition of 1-2 µl ethidium bromide (10 mg/ml) for the detection of DNA bands under UV illumination.

Table 1. List of *Brassica* species and accessions names used in the present study. ND: Not determine.

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 napo*	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

Sequence alignment and phylogenetic analysis of *Brassica CACTA*: The conserved transposase (TNPD) regions (~200 aa) around DDD/E triad motifs of 50 *CACTA* elements from *Brassica* and other plants were collected and aligned in the CLUSTALW implemented in BioEdit (Hall, 1999). The most conserved regions were highlighted by keeping 100% threshold value. The sequence logos were generated from aligned 50 *CACTA* transposase amino acid sequences by online WebLogo (http://weblogo.berkeley.edu/logo.cgi). For the phylogenetic analysis, the aligned amino acid sequences were used to construct the tree in Mega5 (Tamura *et al*., 2011) using Neighbor-Joining method with 1000 bootstraps replicates. The genetic distance was calculated with p-distance method.

Results

***CACTA* identification and structural analyses:** The dot plot comparison of *Brassica* homeologous BAC sequences resulted in the identification of various insertions flanked by 3 bp TSDs, which on detailed structural analysis were characterized as *CACTA* elements. The first autonomous *CACTA* (*BoCACTA1*) was identified by comparing *B. rapa* accession (AC189298.1) against its homeologue *B. oleracea* (EU642504.1). Two other non-autonomous *CACTA* elements were identified by comparing *B. rapa* accession (AC155341.2) against its homeologous *B. oleracea* (AC240089.1). The BLASTN searches of autonomous *BoCACTA1* retrieved several homologues from *B. rapa* and *B. oleracea* with 60-100% homology in their sequences. The *BoCACTA1, BoCACTA2* and *BoCACTA3* elements identified here showed homology to *Bot1-1, Bot1-2,* and *Bot1-3* identified by Alix *et al*. (2008) in *B. oleracea*. Due to their similarity with *Bot1* like elements, these *Brassica CACTA* were considered as members of *Bot1* family. A total of 35 autonomous *CACTA* elements were identified by dot plot analysis and BLASTN searches, of which 19 were from *B. oleracea*, 14 from *B. rapa* and 2 from *B. napus* BACs (Table 2). Seven non-autonomous *CACTA* elements were isolated and characterized from different *Brassica* BACs.

Structural features of *B. oleracea CACTA*: The *BoCACTA* and related homologues displayed typical characteristics of *CACTA* transposons including 3 bp TSDs, TIRs of 15-17 bp, *CACTA* terminal signatures in TIRs and two transposases (TNPD and TNPA). The autonomous elements ranged in sizes from 3 kb to 11 kb. *BoCACTA1* (*Bot1-1*) identified from *B. oleracea* accession (EU642504.1) was 9399 bp large with 3 bp TSDs, 15 bp perfect TIRs (5′-*CACTA*CAAGAAAACA-3′) and displayed both TNPD and TNPA transposases at N and C-terminal ends respectively. A transposase associated domain (TAD) was present towards the N-terminal while two domains of unknown functions (DUF4218, DUF4216) were present towards the C-terminal of TNPD (Fig. 1a). The closest homologues of *BoCACTA1* were *BoCACTA2* and *BoCACTA3*, identified from *B. oleracea* accession (EU642505.1) and (EU642506.1) respectively (Table 2). *BoCACTA2* was 10914 bp with 3 bp TSDs and 15 bp TIRs, while *BoCACTA3* was 11068 bp long including 3 bp TSDs, 15 bp perfect TIRs and several sub-terminal repeats (Table 2). *BoCACTA3* displayed only TNPD transposase, where as TAD domain was located towards N-terminal, while DUF4218 and DUF4216 were located towards C-terminal with an additional domain of unknown function (DUF7241) (Fig. 1b). Interestingly, *BoCACTA2* and *BoCACTA3* captured an ATHILA ORF-1 domain in opposite orientation, which is the integral component of Ty3/gypsy LTR retrotransposons identified in *Arabidopsis thaliana*. Two other *CACTA* (*BoCACTA4* and *BoCACTA5*) were detected in *B. oleracea* accession (EU642505.1) from nucleotide position 21474-29678 and 78098-85744 bp respectively. *BoCACTA4* and *BoCACTA5* were 8205 and 7647 bp long elements with canonical *CACTA* features.

BoCACTA19 (7265 bp) identified from *B. oleracea* accession (EU579455.1) displayed 3 bp TSDs and 15 bp, TNPD transposase, TAD and ATHILA ORF-1 domains (Fig. 1c). The blast analysis retrieved several homologues of this element from *B. oleracea*. Two other *CACTA* designated as *BoCACTA18* and *BoCACTA30* (Fig. 1d) with a size of 10682 and 10728 bp respectively displayed the similar structural features except ATHILA ORF-1 domain is displayed in opposite orientation (Table 2). *BoCACTA21* and *BoCACTA22* were identified as 8210 and 7170 bp large elements encoding both transposase (TNPD and TNPA) with their associated domains. *B. oleracea* accession (AC183496.1) harboured four complete copies of *CACTA* (*BoCACTA30-BoCACTA33*). *BoCACTA30*, the largest (10728 bp) element captured ATHILA ORF-1 domain in it while *BoCACTA31, BoCACTA32* and *BoCACTA33* were 7157, 6075 and 5916 bp large elements (Table 2).

Molecular characterization of *B. rapa CACTA*: The average sizes of *B. rapa CACTA* ranged from 7-8 kb (Table 2). The homologues of *BoCACTA1* (*Bot1*) were also identified and characterized in *B. rapa* genomes. *BrCACTA9* was identified from *B. rapa* accession (AC172883.2) as an insertion from 114211-122180 bp with 3 bp TSDs and 15 bp TIRs (5′-CACTACAAGAAAACA-3′). Using this element as query in BLASTN searches, 14 intact autonomous *CACTA*s were identified from *B. rapa* genome (*BrCACTA6, BrCACTA7, BrCACTA9-BrCACTA17, BrCACTA26, BrCACTA34* and *BrCACTA35*). They have shown similar TIRs as observed in *B. oleracea CACTA* elements. The largest (9393 bp) among the *B. rapa CACTA* was *BrCACTA6* with typical *CACTA* domain patterns (TAD-TNPD-DUF4218-DUF4216-TNPA) (Table 2). *BrCACTA7* (8288 bp) element displayed 3 bp TSDs, 15 bp TIRs (Fig. 1e) similar to *BrCACTA6*. *BrCACTA11* and *BrCACTA16* were 7829 and 5442 bp respectively with canonical *CACTA* protein domain organization (Table 2). A 4952 bp element *BrCACTA17* was identified from *B. rapa* accession (AC189360.2). The smallest (3029 bp) autonomous *CACTA BrCACTA35* displayed a transposase and its associated domain (Fig. 1f).

Table 2. *Brassica CACTA* transposons harbouring in various BAC accessions with their sizes, number of TSDs, TIRs and protein domains organization in BACs.

BAC accessions	Host species	Elements name	Sizes	Position in BACs	TSD	TIR sequences (5′-3′)	Protein domains organization (5′-3′)
EU642504.1	*B. oleracea*	*BoCACTA1*	9399	20580-29972	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
EU642505.1	*B. oleracea*	*BoCACTA2*	10914	44789-55702	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-DUF4271/ATHILA*
EU642506.1	*B. oleracea*	*BoCACTA3*	11068	19777-30844	3	CACTACAAGAAAACA	TNPD-DUF4218-DUF4216/TAD*-ATHILA*
EU642505.1	*B. oleracea*	*BoCACTA4*	8205	21474-29678	3	CACTACAAGAAAACA	TAD-TNPD- DUF4218-DUF4216-TNPA
EU642505.1	*B. oleracea*	*BoCACTA5*	7647	78098-85744	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-TNPA
AC189480.2	*B. rapa*	*BrCACTA6*	9393	87937-97329	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC232490.1	*B. rapa*	*BrCACTA7*	8288	61958-70245	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AJ245479.1	*B. napus*	*BnCACTA8*	8164	44881-53044	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC172883.2	*B. rapa*	*BrCACTA9*	7970	114211-122180	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC189446.2	*B. rapa*	*BrCACTA10*	7861	5462-13322	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216
AC189321.2	*B. rapa*	*BrCACTA11*	7829	92374-100202	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC189341.2	*B. rapa*	*BrCACTA12*	7802	99395-107196	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC189496.2	*B. rapa*	*BrCACTA13*	7779	56849-64627	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC189314.1	*B. rapa*	*BrCACTA14*	7669	21683-29351	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216
AC189655.2	*B. rapa*	*BrCACTA15*	6996	39384-46379	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC189360.2	*B. rapa*	*BrCACTA16*	5442	59073-64514	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC229605.1	*B. rapa*	*BrCACTA17*	4952	83111-88062	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216
AC183492.1	*B. oleracea*	*BoCACTA18*	10682	81000-91686	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218/ATHILA*
EU579455.1	*B. oleracea*	*BoCACTA19*	7265	82206-89482	6	CACTACAAGAAAACA	TAD-TNPD-/ATHILA*
AC183495.1	*B. oleracea*	*BoCACTA20*	9661	104704-114364	3	CACTACAAGAAAACA	TAD-TNPD-/ATHILA*
AC183495.1	*B. oleracea*	*BoCACTA21*	8210	159474-167683	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC183495.1	*B. oleracea*	*BoCACTA22*	7170	237844-245013	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC183493.1	*B. oleracea*	*BoCACTA23*	8072	228710-236781	3	CACTACAAAAAAACA	TAD-TNPD-DUF4218-DUF4216
AC183492.1	*B. oleracea*	*BoCACTA24*	8362	61770-70131	3	CACTACAAGAAAcACA	TAD-TNPD-DUF4218-DUF4216
AC183492.1	*B. oleracea*	*BoCACTA25*	3735	183789-187523	3	CACTACAAGAAAACA	TAD-TNPD
AC172883.2	*B. rapa*	*BrCACTA26*	7970	114211-122180	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC236784.1	*B. napus*	*BnCACTA27*	7192	93542-100733	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216
AC240086.1	*B. oleracea*	*BoCACTA28*	8741	29332-38072	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC240092.1	*B. oleracea*	*BoCACTA29*	9900	32432-42331	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC183496.1	*B. oleracea*	*BoCACTA30*	10728	171084-181811	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-DUF4271/ATHILA*
AC183496.1	*B. oleracea*	*BoCACTA31*	7157	350861-358017	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC183496.1	*B. oleracea*	*BoCACTA32*	6075	302434-308508	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC183496.1	*B. oleracea*	*BoCACTA33*	5916	138717-144632	3	CACTACAAGAAAACA	TAD-TNPD
AC189565.2	*B. rapa*	*BrCACTA34*	5123	57417-62539	3	CACTACAAGAAAACA	TAD-TNPD-DUF4218-DUF4216-TNPA
AC232476.1	*B. rapa*	*BrCACTA35*	3029	93851-96879	3	CACTACAAGAAAACA	TAD-TNPD

TAD: Transposase associated domain. DUF: Domain of unknown function

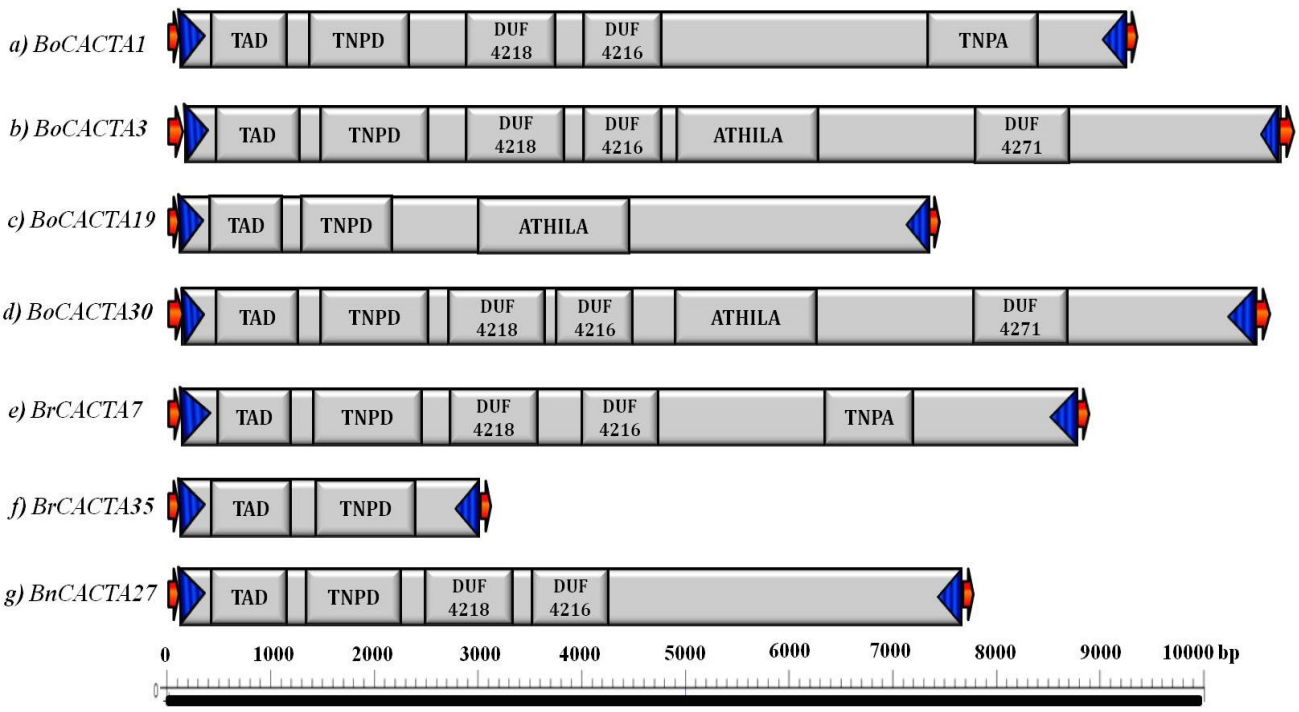

Fig. 1a-g. Schematic representation of various *CACTA* studied in *Brassica*. Red arrows at termini represent TSDs, while blue triangles indicate TIRs. Transposases (TNPD and TNPA), transposase associated domain (TAD), ATHILA-ORF1 domain and domains of unknown functions (DUF4218, DUF4216 and DUF4271) are shown in boxes. The scale below the elements shows their respective sizes in bp.

Identification of *CACTA* elements in *B. napus*: Two complete autonomous *CACTA* and several transposase like sequences were identified from available *B. napus* BACs. *BnCACTA8* was identified from *B. napus* (AJ245479.1) accession with a size of 8164 bp including 3 bp TSDs, 15 bp TIRs, both transposases and associated domains (Table 2). *BnCACTA27* (7192 bp) identified from *B. napus* accession (AC236784.1) displayed 3 bp TSDs, 15 bp TIRs, TNPD transposase and its associated domains (Fig. 1g).

Protein domain organization in *Brassica CACTA*: The autonomous *CACTA* transposons mostly displayed a single transcriptional unit, which generates four to six protein domains (Table 2). TNPD and TNPA (transposase proteins) were detected in most of the elements. The transposase associated domain (TAD) was found located towards N-terminal end of TNPD. The exact function of TAD is not known but it is the accessory component of TNPD transposase. Two domains named DUF4218 and DUF4216 were identified towards the C-terminus. The domain organization of autonomous *CACTA* from *Brassica* and other plants revealed two major patterns. The first pattern was displayed by majority of *Brassica CACTA* as 5′-TAD-TNPD-DUF4218-DUF4216-TNPA-3′. The second pattern of protein domain organizations was 5′-TAD^{+}-TNPD^{+}-DUF4218^{+}-DUF4216^{+}-[ATHILA-ORF1^{-}]-TNPA^{+}-3′, where signs + and – indicate plus and minus orientations (Table 2).

***Brassica CACTA* captures ATHILA ORF-1 domain:** *Brassica CACTA* transposons captured an ATHILA ORF-1 domain in their coding regions, which is the integral part of *Arabidopsis thaliana* Ty3/gypsy LTR retrotransposons. The *B. oleracea CACTA* showed homology in ~1200-1280 bp (~400-428 aa) region of ATHILA ORF-1 domain from *A. thaliana* Gypsy retrotransposon. The ATHILA ORF-1 domain was found inserted in *BoCACTA2, BoCACTA3, BoCACTA18, BoCACTA19, BoCACTA20* and *BoCACTA30* (Table 2), which increases the sizes of these elements. In general, a 3.1 kb insertion was detected in *Brassica CACTA,* with ~1.2 kb region homologous to ATHILA ORF-1 domain.

Characterization of non-autonomous *CACTA*: Non-autonomous *Brassica CACTAs* were smaller in sizes ranging from 1.2 kb to 3.2 kb (Fig. 2). *Bo-N-CACTA1* (3265 bp) was identified from *B. oleracea* accession (AC240092.1) from nucleotide position 48182-51446 bp with 3 bp TSDs and 15 bp TIRs (5′-CACTGGTGGAGAAACC-3′). *Br-N-CACTA2* (2559 bp) was identified from *B. rapa* accession (AC155342.2) from 58153-60711 bp within BAC sequence. The 300 bp terminal regions were used to locate its autonomous *CACTA*, where we found *BrCACTA6* and related homologues as its progenitors. *Bo-N-CACTA3* and *Bo-N-CACTA4* were 2662 and 2773 bp large elements with 3 bp TSDs and 15 bp TIRs (Fig. 2). The comparison of *B. rapa* accessions (AC155341.2) against *B. rapa* (AC189489.2) resulted in the identification of *Br-N-CACTA5* and *Br-N-CACTA6* with a size of 1419 and 1288 bp respectively. *Br-N-CACTA7* (1288 bp) was identified as a homologue of *Br-N-CACTA6* residing in *B. rapa accession* (AC241034.1) (Fig. 2).

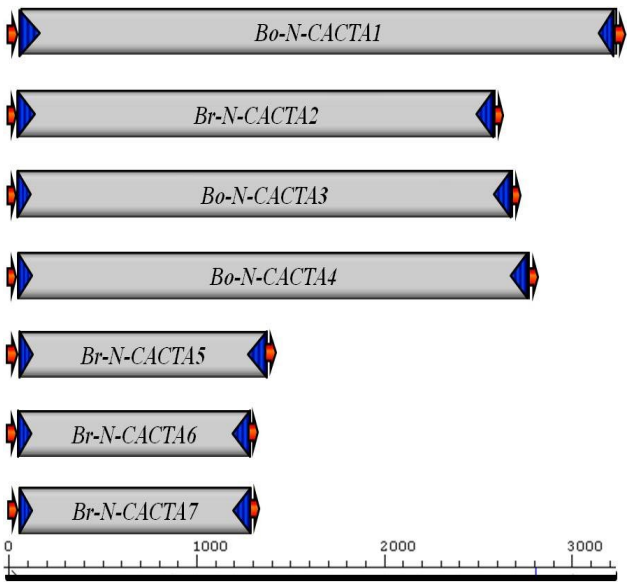

Fig. 2. Schematic representation of non-autonomous *CACTA* elements studied in *Brassica*. TSDs are shown by red arrows, while blue triangles indicate TIRs. The elements have not shown any protein coding domains like transposase or any other protein. The scale below the elements shows their sizes in bp.

PCR amplification of *CACTA* transposase (TNPD): To amplify *Brassica CACTA* transposase, degenerate primers pair BoCACTAF and BoCACTAR (see material and methods) was designed from the conserved DDD/E region of transposase (TNPD). PCR was carried out to amplify the 580 bp (~190 aa) of DDD/E domain region of transposase. *Brassica CACTA* transposase was successfully amplified from all the 40 diploid and polyploids *Brassica* lines (Fig. 3a) suggesting its high diversity among *Brassica* species. A 580 bp strong band of transposase was amplified from all A-genome *B. rapa*, B-genome *B. nigra*, C-genome *B. oleracea*, allotetraploid *B. juncea*, *B. napus, B. carinata* and all synthetic hexaploid *Brassicas*. The amplification of *CACTA* in all *Brassica* species indicated its diversity, distribution, their ancient nature and suggested their amplification before the separation of *B. nigra* and *Brassica rapa/B. oleracea*.

PCR amplification of ATHILA ORF-1 in *Brassica*: To investigate, whether ATHILA ORF-1 was only captured by *B. oleracea CACTA* or *B. rapa CACTA* also harboured it, the primers BoATHILAF and BoATHILAR (see material and methods) were designed from the ATHILA ORF-1 domain. Of the 40 *Brassica* diploids and polyploids accessions/genotypes (Table 1), a 1 kb ATHILA ORF-1 was amplified from 28 accessions (Fig. 3b) indicating its presence in most of the *Brassica* genomes. A weak band of ~1 kb size was amplified from *B. rapa* (Pak Choy, San Yue Man, Vertus, Suttons) genotypes. All the three *B. nigra* genotypes failed to amplify ATHILA ORF-1 domain. All the six *B. oleracea* genotypes amplified the 1 kb band of ATHILA ORF-1. Of the nine *B. juncea*, NARC-II, Kai Choy and W3 amplified the products. Strong bands were amplified from all the six *B. napus* genotypes.

Phylogenetic analysis of *Brassica CACTA* transposase (TNPD): The alignment of 35 *Brassica* and 15 other plants transposases (TNPD) was performed by CLUSTALW available in BioEdit program. The 35 *Brassica* and 5 *A. thaliana* transposases were retrieved from NCBI, while other 10 transposases from various plants were collected from Repbase database (Jurka *et al.*, 2005). The alignment of 50 transposases allowed the identification of motifs essential for the transposition, which were mostly perfect but in few cases were interrupted by stop codons, frame shift mutations or lacking the translation initiation codons. The highly conserved catalytic triad motifs $DD_{39}D$ and $DD_{32}E$ were present in all the transposases with few other conserved amino acid motifs (Fig. 4). The amino acid residues around the $DD_{39}D$ triad and between the second and third aspartic acid residue ($D_{39}D$) were most conserved among all plant transposases.

To gain into the evolutionary relationship of *Brassica* and other plant *CACTA*, phylogenetic tree was constructed based on amino acid sequences of 50 conserved transposases (TNPD; ~ 200 aa residues), which clustered them into two major lineages (Fig. 5). One lineage was represented by 7 monocot and dicot transposases, while other lineage clustered other 43 Brassicaceae related *CACTA* except *Chester-1* of *A. thaliana*. The first lineage represented by 7 transposases further splits into two clades. The first clade (ENSPM) was represented by the grass family members as *EnSpm10_TM* from *Triticum monococcum*, *EnSpm10_OS* from *Oryza sativa* and *EnSpm10_ZM* from *Zea mays*. In the second clade (CHESTER1), *Chester-1* of *A. thaliana*, *EnSpm-13* of *Vitis vinifera, TGM5* of *Glycine max, TDC1* of *Daucus carota* and *PSL* of *Petunia hybrida* clustered together. The second lineage was represented by 43 *CACTA* transposases from *Arabidopsis* and *Brassica* (Fig. 5) suggesting their monophyletic origin from a common ancestor before the separation of two genera around 20 Mya. This lineage further resolved into 2 clades with 1 (*ATCACTA3*) and 42 elements in each clade. The second clade with 42 elements was further resolved into 8 families with 6, 2, 4, 2, 1, 1, 12, 12 elements in each family (Fig. 5). The *ENSPM_B. rapa* clustered with other 5 elements from *B. oleracea* in first family. Second family clustered 2 elements, while the four *A. thaliana* elements (*ATCACTA1, ATCACTA2, ATCACTA4, ATCACTA5*) clustered together in third family and constituted a sister family with *B. oleracea* elements. The 9 *B. oleracea* along with 2 *B. rapa* (*BrCACTA9, BrCACTA11*) and one *B. napus* transposase clustered in seventh family, while family eight is comprised of 12 *B. rapa* mediated transposases. The evolutionary analysis suggested that *Brassica CACTA* transposases are not only conserved in diploid *Brassicas* but actively proliferating in allotetraploid *Brassicas* (*B. juncea, B. napus, B. carinata*) and sister genera *Arabidopsis*.

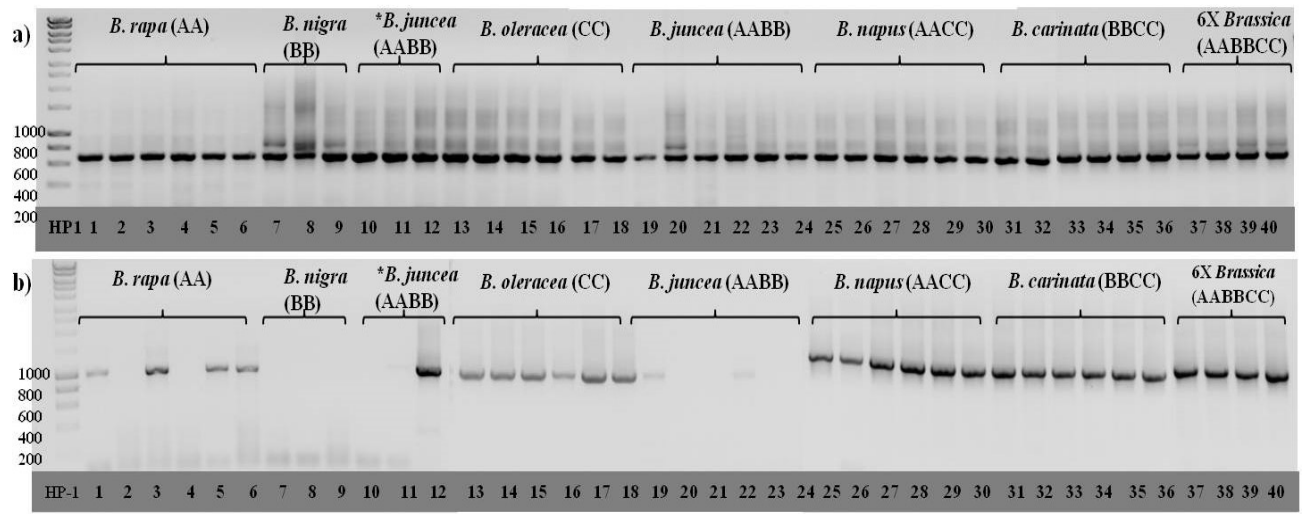

Fig. 3. PCR amplification of a) *CACTA* TNPD-transposase from 40 *Brassica* lines. The 580 bp bands amplify the *CACTA* transposase from all 40 *Brassica* genomes. b) BoATHILA ORF-1 domain: the 1000 bp band shows the presence of this domain in *Brassica* but in contrast to the *CACTA* transposase, it is not present in all accessions of *B. rapa, B. nigra* and *B. juncea*. Numbers below the lanes identify each genotype listed in table 1 and ladders indicate sizes in bp.

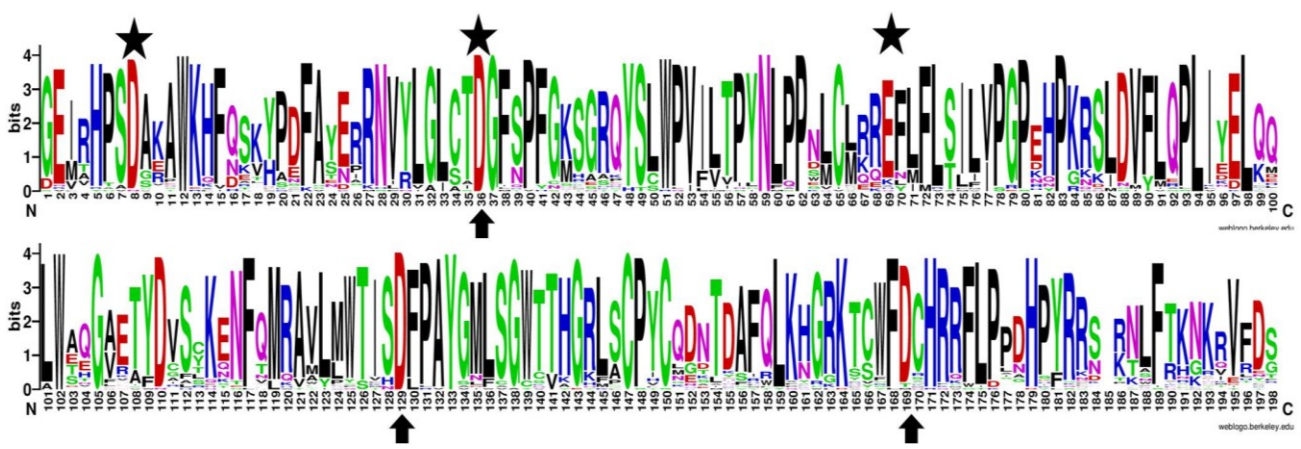

Fig. 4. Weblogo based on 50 *Brassica* and other plants *CACTA* transposases (200aa). The transposase sequences are highly conserved in *Brassica* and *Arabidopsis*. The height of nucleotides (0 to 4) is proportional to their conservation. The DDE or DDD traid motifs are represented by stars and arrows respectively.

Discussion

Transposable elements as ubiquitous components of major eukaryotic genomes played a major role in genome duplications and plasticity. Several computational programs and tools were recently developed for the identification of TEs, but it is still challenging to proper identify and characterize them due to their several structural modifications or deformations (Gao *et al*., 2014; Nouroz *et al*., 2015a). The present study involved identification of DNA-mediated *CACTA* transposons by comparing *Brassica* BACs in JDotter program, a highly efficient program which identify the small insertions in one or the other genomic sequence. In the present study, 35 autonomous (Table 2) and 7 non-autonomous *CACTA* elements and their several analogues were detected proliferating in the *Brassica* genome. It was found that the first identified element showed 100% homology to the *Bot1* family (Alix *et al.,* 2008), due to the reason these *Brassica CACTA* were placed in *Bot1* family. The *Brassica Bot1* family was also investigated in *Arabidopsis*, where ~110 copies of *Bot1*-like transposase were isolated suggesting their abundance and proliferation in *Arabidopsis* genome and their origin from a common ancestor. This was confirmed by computational based comparative analysis of *Brassica* and *Arabidopsis*, indicating that both share the same collection of TEs but in varied proportions, the number being greater in *B. oleracea* due to three fold larger genome than *Arabidopsis* (Zhang *et al.*, 2004). The present study indicated that *CACTA* elements from A and C-genome *Brassica* have shown high homology in their sequences especially in TIRs (98-100%). The homology within *CACTA* sequences remained consistent among *Brassica* and *Arabidopsis*. This is in accordance to the investigations of Zhang & Wessler (2004), who analyzed the evolutionary relationship of *CACTA* transposons in *Brassica* and *Arabidopsis* and showed high intra-family homology of *B. oleracea CACTA* with a close relation to *Arabidopsis*. The molecular analysis of *CACTA* investigated in present study revealed that they encode two transposases (TNPD, TNPA) and 1-3 additional proteins. Such additional proteins were also observed in *Casper* family in *Triticeae* (Wicker *et al.*, 2003).

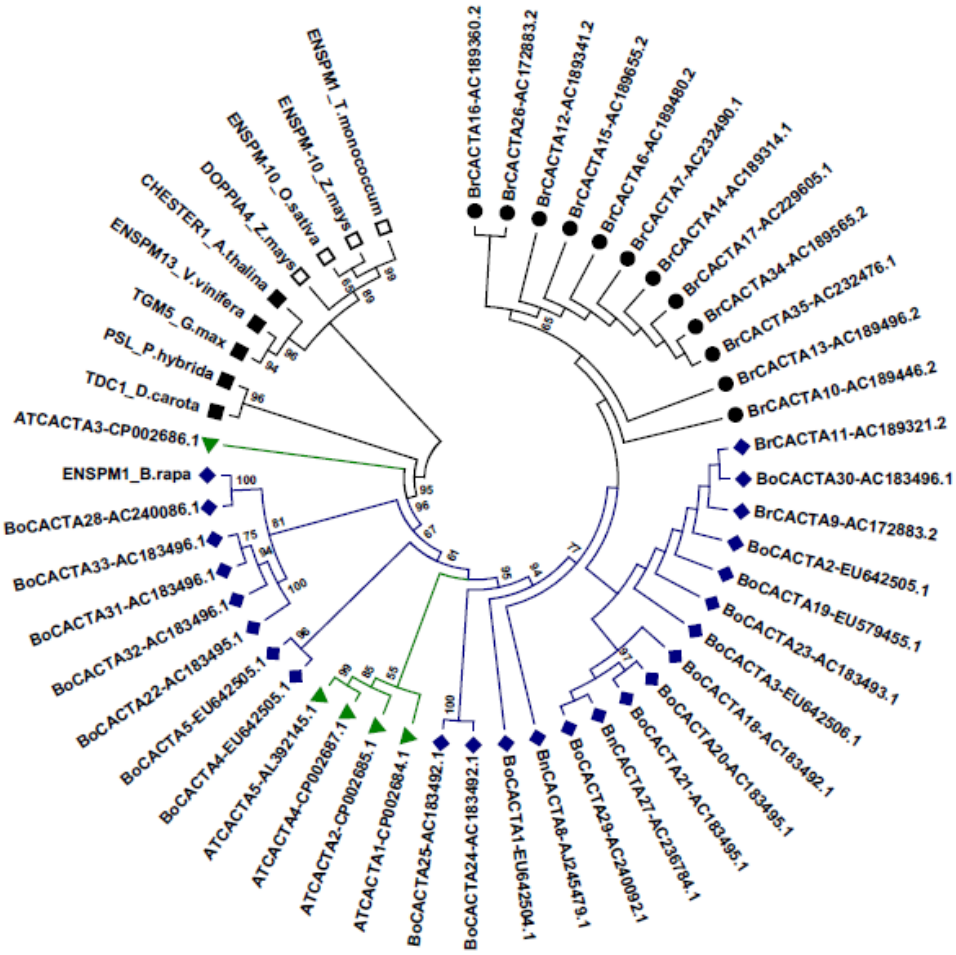

Fig. 5. Neighbor-Joining tree showing relationship of *CACTA* family TNPD-transposase (Transposase-21). The phylogenetic tree of *Brassica CACTA* based on protein sequence of transposases was constructed by the Neighbor-Joining method with 1000 bootstrap replicates using the Mega5 program. The bootstrap support (%) is shown near the nodes. Various families are represented by different open and filled shapes/colours and branches. The names of the elements are followed by the BAC accession names from which they were identified.

The identification of several autonomous *CACTA* and their non-autonomous partners revealed their abundance in *Brassica* genome. Of the 42 *Brassica CACTA* characterized, 35 were autonomous and 7 were non-autonomous. All the 35 elements encoded TNPD transposase but 15 *Brassica CACTA* lack the TNPA transposase. Although few of these autonomous *CACTA* have frame shift mutations or in-frame stop codons within their coding regions, but all those elements were considered intact elements due to presence of TSDs, TIRs and transposase. Several non-autonomous *CACTA* were observed in *Brassica*, but their mechanism of transposition is not clear. It is more likely that they utilize transposase of autonomous *CACTA* elements residing nearby. The diversity and abundance of *Bot1* was investigated in *B. oleracea* genome, where large sized (9.3-11 kb) *Bot1* elements played a vital role in *B. rapa* and *B. oleracea* genome divergence by proliferating in *B. oleracea* (Alix *et al.*, 2008). These results are in parallels with the results in Solanaceae, where *CACTA* diversity was investigated in several plants such as *Solanum tuberosum*, *Nicotiana tabacum* and *Datura stramonium* (Proels and Roitsch, 2006). The soybean genome harbours several *CACTA* elements in their genomes, where nine *CACTA* elements designated as *Tgm1*, *Tgm2*, *Tgm3*, *Tgm4*, *Tgm5*, *Tgm6*, *Tgm7*, *Tgm-Express1*, and *Tgmt** have been reported (Zabala & Vodkin, 2008). Among various DNA superfamilies of TEs, the maximum copy numbers (348) of *CACTA* elements were observed in *Phaseolus vulgaris* (Gao *et al*., 2014). The monocot genome is also a hotspot for *CACTA* proliferation (Wicker *et al*., 2003).

The number and conserved pattern of TIRs specify a DNA transposon superfamily. The TIRs in *Brassica CACTA* were 15 bp and were highly conserved (5′-*CACTA*CAAGAAAACA-3′) with the exception of 1 autonomous (*BoCACTA24*) and a non-autonomous (*Br-N-CACTA6*) element, where one or two additional nucleotides were detected upstream to the 3′-termini of TIRs. Similar *CACTA* TIRs were reported from *Brassica* (Alix *et al.*, 2008). The TIRs of *Brassica CACTA* were compared with TIRs of other plants *CACTA* collected from Repbase database (Jurka *et al.*, 2005). The *BRENSPM1* element from *B. rapa* also possess similar 15 bp TIRs (5′-*CACTA*CAAGAAAACA-3′). The closest genera (*A. thaliana*) of *Brassica* have shown more or less similar TIRs such as *Chester-1*. In contrast the element *CAC1* from *Arabidopsis thaliana* generates the shortest TIRs (5′-CACTACAA-3′), which are completely similar

to 5′ termini of TIRs. The similarity of TIRs in *Brassica* and *Arabidopsis* suggests their common origin from the same ancestor before separation of both genera; however the *CACTA* superfamily is evolutionarily much older (Buchmann *et al*., 2014). The *PSL* element from *P. hybrida*, *EnSpm-13_VV* element from *V. vinifera* and *EnSpm_MT* elements from *M. truncatula* displayed 12, 13 and 14 bp TIRs respectively. The TIRs of Soybean *Tgm1* showed 30 bp TIRs with homology in first 14 nucleotides (Xu *et al.*, 2010). The overall review of plant *CACTA* revealed that the '*CACTA*' signature motif is the most conserved in all *CACTA* elements.

The phylogenetic analysis of the present study based on alignment of *Brassica, Arabidopsis* and other monocotyledonous plant *CACTA* transposases resulted in clustering of two lineages; first lineage is mostly comprised of monocotyledonous transposases along with few dicotyledonous transposases, while second lineage clustered transposases from *Brassica, Arabidopsis* and other dicotyledonous plants. These results revealed that *CACTA* lineages diverged before the divergence of monocotyledonous and dicotyledonous plants. Such divergence of *CACTA* transposase before the divergence of monocots and dicots was observed by aligning 64 transposases from various plants. The presence of mixed clades and the close relationship of these clades from both groups revealed the ancient divergence of *CACTA* superfamily (Buchmann *et al*., 2014). On the other hand high homology seen in *Brassica CACTA* transposases from various species has supported their monophyletic origin.

Conclusion

Repetitive DNA due to their importance is the centre of focus now days. Our identification and characterization of *Brassica CACTA* transposons by dot plot comparison of *Brassica* BACs and database mining gives an insight into the structural and evolutionary dynamics of *Brassica CACTA* in detail. The results described the variations in structures and sizes of *CACTA* elements especially in *Brassica* and its allied genera *Arabidopsis* with apparent analysis of *CACTA* belonging to few monocotyledonous plants. Present study indicates a high homology among *Brassica/Arabidopsis CACTA* transposases and slight variation among *CACTA* transposases of monocotyledons and dicotyledonous plants indicating their ancient divergence. The clustering of transposases from various *Brassica* species into same or sister families revealed their common ancestry before their divergence around 17 Mya. The identification of these *CACTA* elements could be helpful in the discovery of active transposons, which can be used for transposon tagging system as utilized previously in case of *Ac/Ds* or *En/Spm* elements.

Acknowledgements

The financial support for this work was provided by Hazara University Mansehra and Higher Education Commission, Pakistan. The work was conducted in highly equipped laboratories of Department of Biology, University of Leicester, UK. The seeds/DNA material was provided by Warwick Research Centre, UK and Dr. Xian Hong Ge, University of Huazhong Agricultural University, Wuhan, China.

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.

Buchmann, J.P., A. Loytynoja, T. Wicker and A. Schulman. 2014. Analysis of *CACTA* transposases reveals intron loss as major factor influencing intron/exon structure in monocotyledonous and eudicotyledonous hosts. *Mobile DNA*, 5: 24.

Christopher, G.L., J.R. Andrew, A.C.L. Geraldine, J.H. Clare, B. Jacqueline, B. Gary, C.S. German and E. David. 2005. *Brassica* ASTRA: an integrated database for *Brassica* genomic research. *Nucleic Acids Res*., 33(2): 656-659.

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

Gao, D., B. Abernathy, D. Rohksar, J. Schmutz and S.A. Jakson. 2014. Annotation and sequence diversity of transposable elements in common bean (*Phaseolus vulgaris*). *Front. Plant Sci.,* 5: 1-9.

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.

Gierl, A. 1996. The *En/Spm* transposable element of maize. *Curr. Top. Microbiol. Immunol.,* 204: 145-159.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.*, 41: 95-98.

Jurka, J., V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany and J. Walichiewicz. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.,* 110: 462-467.

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

Koch, M.A., B. Haubold and T. Mitchell-Olds. 2000. Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (*Brassicaceae*). *Mol. Biol. Evol.,* 17: 1483-1498.

Lee, J.K., S.J. Kwon, K.C. Park and N.S. Kim. 2005. *Isaac-CACTA* transposons: new genetic markers in maize and sorghum. *Genome,* 48: 455-460.

Miura, A., S. Yonebayashi, K. Watanabe, T. Toyama, H. Shimada and T. Kakutani. 2001. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature,* 411: 212-214.

Monteiro, A. and T. Lunn. 1999. Trends and perspectives of perspectives of vegetable *Brassica* Breeding World-wide. WCHR- World Conference on Horticulture Research ISHS. *Acta Horticulture,* 495.

Nouroz, F., S. Noreen and J.S. Heslop-Harrison. 2015b. Molecular characterization and diversity of a novel non-autonomous *mutator-like* transposon family in *Brassica. Pak. J. Bot*., 47(4): 1367-1375.

Nouroz, F., S. Noreen and J.S. Heslop-Harrison. 2015c. Identification, characterization and diversification of nonautonomous hAT transposons and unknown insertions in *Brassica. Genes Genom*., 37: 945-958.

Nouroz, F., S. Noreen and J.S. Heslop-Harrison. 2016. Characterization and diversity of novel *PIF/Harbinger* DNA transposons in *Brassica* genomes. *Pak. J. Bot.,* 48(1): 167-178.

Nouroz, F., S. Noreen, J.S. Heslop-Harrison. 2015a. Identification and characterization of LTR Retrotransposons in *Brassica*. *Turk. J. Biol*., 39: 740-757.

Ostergaard, L. and G.J. King. 2008. Standardized gene nomenclature for the *Brassica* genus. *Plant Methods,* 4: 10.

Pereira, A., H. Cuypers, A. Gierl, Z. Schwarz-Sommer and H. Saedler. 1986. Molecular analysis of the *En/Spm* transposable element system of *Zea mays*. *Embo. J.,* 5: 835-841.

Proels, R.K. and T. Roitsch. 2006. Cloning of a *CACTA* transposon-like insertion in intron I of tomato invertase Lin5 gene and identification of transposase-like sequences of *Solanaceae* species. *J. Plant Physiol.,* 163: 562-569.

Roccaro, M., Y. Li, H. Sommer and H. Saedler. 2007. *ROSINA* (RSI) is part of a *CACTA* transposable element, TamRSI, and links flower development to transposon activity. *Mol. Genet. Genomics,* 278: 243-254.

Sergeeva, E.M., E.A. Salina, I.G. Adonina and B. Chalhoub. 2010. Evolutionary analysis of the *CACTA* DNA-transposon Caspar across wheat species using sequence comparison and in situ hybridization. *Mol. Genet. Genomics,* 284: 11-23.

Sidra, I., Farhatullah, S. Shah, M. Kanwal, L. Fayyaz and M. Afzal. 2014. Genetic variability and heritability studies in indigenous *Brassica rapa* accessions. *Pak. J. Bot*., 46(2): 609-612.

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: 1-10.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum parsimony methods. *Mol. Biol. Evol.*, 28: 2731-2739.

Tian, P.F. 2006. Progress in plant *CACTA* elements. *Yi Chuan Xue Bao,* 33: 765-774.

Trentmann, S.M., H. Saedler and A. Gierl. 1993. The transposable element *En/Spm*-encoded TNPA protein contains a DNA binding and a dimerization domain. *Mol. Gen. Genet.,* 238: 201-208.

Wang, G.D., P.F. Tian and Z.K. Cheng. 2003. Genomic characterization of *Rim2/Hipa* elements reveals a *CACTA*-like transposon superfamily with unique features in the rice genome. *Mol. Genet. Genomics,* 270: 234-42.

Wicker, T., F. Sabot, A. Hua-Van and J.L. Bennetzen. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.,* 8(12): 973-982.

Wicker, T., R. Guyot, N. Yahiaoui and B. Keller. 2003. *CACTA* transposons in *Triticeae*. A diverse family of high-copy repetitive elements. *Plant Physiol,* 132: 52-63.

Xu, M., H.K. Brar, S. Grosic, R.G. Palmer and M.K. Bhattacharyya. 2010. Excision of an active *CACTA*-like transposable element from DFR2 causes variegated flowers in soybean [*Glycine max* (L.) Merr.]. *Genetics,* 184: 53-63.

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

Zabala, G. and L. Vodkin. 2008. A putative autonomous 20.5 kb-*CACTA* transposon insertion in an F3'H allele identifies a new *CACTA* transposon subfamily in *Glycine max*. *BMC Plant Biol.,* 8: 124.

Zhang, X. and S.R. Wessler. 2004. Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *PNAS*, 15: 5589-5594.

(Received for publication 10 January 2016)