Pak. J. Bot., 48(1): 167-178, 2016.

CHARACTERIZATION AND DIVERSITY OF NOVEL *PIF/HARBINGER* DNA TRANSPOSONS IN *BRASSICA* GENOMES

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

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

Abstract

Among DNA transposons, *PIF/Harbinger* is most recently identified superfamily characterized by 3 bp target site duplications (TSDs), flanked by 14-45 bp terminal inverted repeats (TIRs) and displaying DDD or DDE domain displaying transposase. Their autonomous elements contain two open reading frames, ORF1 and ORF2 encoding superfamily specific transposase and DNA-binding domain. *Harbinger* DNA transposons are recently identified in few plants. In present study, computational and molecular approaches were used for the identification of 8 *Harbinger* transposons, of which only 2 were complete with putative transposase, while rest 6 lack transposase and are considered as defective or non-autonomous elements. They ranged in size from 0.5-4 kb with 3 bp TSDs, 15-42 bp TIRs and internal AT rich regions. The PCR amplification of *Brassica Harbinger* transposase revealed diversity and ancient nature of these elements. The amplification polymorphism of some non-autonomous *Harbingers* showed species specific distribution. Phylogenetic analyses of transposase clustered them into two clades (monocot and dicot) and five sub-clades. The *Brassica*, *Arabidopsis* and *Malus* transposase clustered into genera specific sub-clades; although a lot of homology in transposase was observed. The multiple sequence alignment of *Brassica* and related transposase showed homology in five conserved blocks. The $DD_{35}E$ triad and sequences showed similarity to already known *Pong*-like or *Arabidopsis ATISI12 Harbinger* transposase in contrast to other transposase having $DD_{47}E$ or $DD_{48}E$ motifs. The present study will be helpful in the characterization of *Harbingers*, their structural diversity in related genera and *Harbinger* based molecular markers for varietal/lines identifications.

Key words: DNA transposons, *Harbinger*, *Brassica*, SANT, Transposase, DDE motif.

Introduction

Transposable elements (TEs); the mobile DNA elements are disperse repetitive sequences of almost all plant-animal-fungal genomes with diversity in various important agricultural crops like maize, wheat, barley, rye sugar beet and *Brassica*, where 50-90% genome is composed of TEs (Kubis *et al*., 1998; Wicker & Keller, 2007; Kapitonov & Jurka, 2008; Nouroz *et al*., 2015a). The larger genomes are made up of abundant tandem repetitive sequences and TEs, which compose a major proportion of DNA, sometimes representing more than half of DNA (Heslop-Harrison & Schwarzacher, 2011). Based on their transposition mechanism, they are classified into Retrotransposons and DNA transposons (Finnegan, 1989). Presence or absence of protein domains (reverse transcriptase/transposase) required for their transposition further divide them into autonomous (complete) and non-autonomous (defective) elements respectively. The DNA transposons are composed of several superfamilies; the major among plant genomes are *Tc1/ Mariner*, *hAT*, *CACTA*, *Mutator* and *PIF/Harbinger* (Wicker *et al*., 2007; Kapitonov & Jurka, 2008, Nouroz, 2012).

The *PIF/Harbinger* is most recently identified superfamily of DNA transposons characterized by 3 bp target site duplications (TSDs), flanked by 14-25 or up to 50 bp terminal inverted repeats (TIRs) and displaying a DDD or DDE domain containing transposase showing similarity to bacterial IS5 insertion sequence (Kapitonov & Jurka, 2004; Zhang *et al*., 2001, 2004). *PIF/Harbinger* gained its name by genetic discovery of two founder elements; *PIF* from *Zea mays* and *Harbinger* from *Arabidopsis thaliana*. The diverse *PIF/Harbinger* elements are easily distinguishable into two subgroups, named *PIF* and *Pong*, which are distributed in several plant genomes (Kapitanov & Jurka, 1999; Zhang *et al.*, 2004). *Harbinger* is highly diverse superfamily of DNA transposons with members distributed among protists, insects, worms, vertebrates and plants. Their autonomous elements contain two open reading frames, ORF1 and ORF2, in which superfamily specific DDE catalytic transposase and DNA-binding domains are encoded. The DNA binding domain is characterized by having different conserved motifs as SANT/myb/trihelix (~70 aa) (Kapitonov & Jurka, 2004; Markova, 2014). Generally, *Harbinger* are flanked by TAA/TTA target site duplications, but some families generate other TSDs, as CAG target sites observed in Zebra fish *Harbinger2-3_DR* (Kapitonov & Jurka, 2004). The phylogenetic studies based on *Harbinger* transposases suggest their horizontal transfer. As, the transposases from the *Arabidopsis* and maize *Harbinger* and *PIF* elements are more similar to diatom *Harbinger1-2_TP* transposase as compared to their closely related rice *Pong* and *Arabidopsis ATIS112A.* The *PIF* and *Harbinger* were considered as two separate superfamilies prior to 2001, which merged to a single superfamily due to high similarities between the elements (Jurka & Kapitonov, 2001; Kapitonov & Jurka, 2004; Markova, 2014). The non-autonomous *PIF/Harbinger* elements are named miniature inverted repeat transposable elements (MITEs) belonging to *Tourist* family by several authors and are common in several plants like *Brassica* (Nouroz *et al*., 2015b). The *Tourist* MITEs are short elements (<500 bp) displaying TIRs but lacked internal transposase coding domains. *Tourist* MITE *mPING* with several active copies

proliferating in rice (Jiang *et al*., 2003) and yeast genomes (Hancock *et al*., 2010) were identified.

In the previous years, *PIF/Harbingers* were identified from few plants indicating their active proliferation in their genomes. Twenty two putative autonomous and 67 non-autonomous *PIF/Harbinger* elements were identified from *Medicago truncatula*, further divided into five families; three previously identified and two newly identified families (Grzebelus *et al*., 2007). Two elements *DcMaster1* (2.5 kb) and *DcMaster-a* (4.4 kb) were identified from carrot genomes with several other homologous copies dispersed in genome (Grzebelus *et al*., 2006, 2009). Another *PIF* like transposase containing *Harbinger Boto* was identified from fungus *Moniliophthora perniciosa*. *Boto* showed sequence similarity with known *PIF* like elements from plant genomes and displayed the $DD_{48}E$ transposase domain as identified in other plant genomes (Pereira *et al*., 2013). Around 139 *PIF* related sequences were identified from 44 Bamboo species indicating their abundance and diversity in Bambusoideae subfamily (Zhou *et al*., 2010, 2012). The detailed evolutionary analysis of *PIF* and *Pong* transposase among Triticeae genomes revealed their wide distribution in Triticeae tribe sharing several structural features (Markova, 2014).

With the advancement in sequencing, the annotation and genomic diversity of various TEs is important. However, the information about the TEs especially DNA transposons in plant genomes is very limited. The *PIF/Harbingers* are identified and characterized from very few plant genomes. The present study involved the characterization of novel *Harbinger* transposons and their diversity in economically important *Brassica* genomes with special emphasis on their structural diversity, distributions and evolutionary scenario.

Material and Methods

Plant material for *Brassica*: Standard CTAB method (Doyle & Doyle, 1990) was adopted for DNA extraction from 40 *Brassica* accessions/cultivars (Table 1). Seeds from 32 *Brassica* accessions were a gift from Warwick Research Institute (WRI), Warwick, UK. Two *B. juncea* (NARC-1, NARC-II) and one *B. carinata* (NARC-PK) accession was sent from National Agriculture and Research Center (NARC), Islamabad, Pakistan. *B. juncea* (NATCO) seeds were purchased from Asian store at Leicester, while DNA from 4 synthetic allohexaploids (2n=6x) (Ge *et al.*, 2009) was provided by Xian Hong Ge (University of Wuhan, China). The seeds were grown in a green house at Department of Biology, University of Leicester, UK.

Computational analysis for characterization of *Brassica Harbinger*: Dot plot analysis was performed for *de novo* identification of *Brassica Harbinger* elements. Homeologous *B. rapa* (AA) and *B. 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 (Altschul *et al.*, 1997; Altschul *et al.*, 2009). The elements were characterized on the basis of their structural hallmarks (TSDs and TIRs, transposase and associated domains) into their respective superfamily and family. The names to the *Harbingers* were given according to the recommendations of Capy (2005) for naming TEs, such as ***BoHARB1***, where ***B*** stands for genus *Brassica*, ***o/r*** represents *oleracea/rapa*, ***HARB*** indicate transposons superfamily and number **1** indicate the family. The letter **N** used before the superfamily name indicates non-autonomous elements as ***BoN-HARB1***.

Table 1. List of *Brassica* species and accessions with their accessions and crop names. 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*	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	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

PCR amplification of *Brassica Harbinger*: The degenerative primers were designed around DDE triad motif from the conserved RT regions by using Primer3 (http://frodo.wi.mit.edu/primer3/). In non-autonomous *Harbingers* primers were designed from N-terminal and C-terminal regions. Total 50 ng *Brassica* genomic DNA was used for PCR amplification in a 15 µl reaction mixture containing 2 µl PCR buffer (KAPPA, UK), 1.0 mM additional $MgCl_2$, 1U KAPPA Taq DNA polymerase (KAPPA, UK), 200-250 mM dNTPs and 0.75 µl (10 pmoles) of each primer. The thermal cycling conditions were adjusted as: 3 min denaturation at 94˚C; 35 cycles of 45 sec denaturation at 94˚C, 45 sec annealing at 52-64˚C (primer dependent), 1 min extension at 72˚C and final 3 min extension at 72˚C. PCR products were separated by electrophoresis in 1% Agarose gel according to the standard protocols. Gels were stained with 1-2 µl ethidium bromide (10 mg/ml) to detect DNA bands under UV illumination.

Sequence alignment and phylogenetic analysis of *Brassica Harbinger*: The GC and AT contents of the *Harbinger* were calculated using the online "GC Calculator" (http://www.genomicsplace.com/gc_calc.html). To detect domains in sequences, they were blasted against the 'Conserved Domain Database' available in NCBI. The Weblogos were generated by online website of Weblogo (http://weblogo.berkeley.edu/logo.cgi). For the phylogenetic analysis, the conserved DDE transposase domain (~200 aa) were aligned in the CLUSTALW implemented in BioEdit program (Hall, 1999). Tree was generated in Mega5 program (Tamura *et al*., 2011) using Neighbour-Joining method with 1000 bootstraps replicates.

Results

Identification and structural analysis of *Brassica Harbingers*: The first identified *Harbinger* '*BoHARB1*' was identified in *B. oleracea* BAC accession 'AC240081.1' inserted at position 5984-9826 bp. The element displayed a size of 3843 bp, generates a typical *Harbinger*-like TAA target site duplication and flanked by 42 bp TIRs (Fig. 1; Table 2). The *BoHARB1* is highly AT rich (60%), with high AT rich region (75%) in the first 350 bp immediately after the 5´ TIR. The detailed analysis of structural domains revealed that element exhibits SANT domain only, while lack transposase (TNP), due to which it is considered as a defective *Harbinger*. Another *Harbinger*-like insertion *BoHARB2* was identified from *B. oleracea* accession 'AC240081.1' from position 53192-56946 bp. The 3755 bp insertion exhibit the structural features of *Harbinger* displaying TTA TSDs and 15 bp TIRs (Fig. 1; Table 2). The element is AT rich (63%) with many small poly A/T sequences dispersed within the molecule. The molecular organization of *BoHARB2* displayed the encoding of two protein domains as thioredoxin (TRX) and ATP11, located at sub-terminal region of C-terminal (3´) end of element.

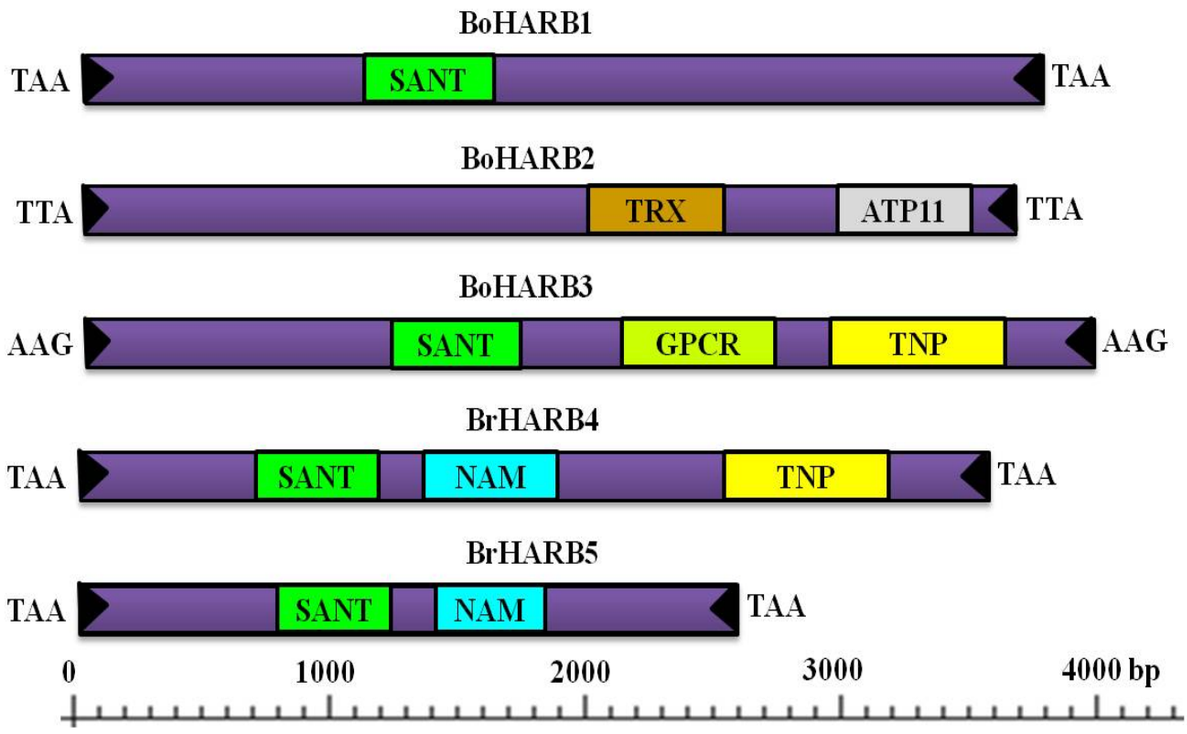


Fig. 1. Schematic representation of *Brassica Harbingers*. The 3 bp at termini represent TSDs, while black triangles indicate TIRs. The transposase (TNP), SANT, NAM and other domains are represented with different colours. The protein domains were identified by screening these sequences against known proteins in the conserved domain database (CDD). The scale below shows sizes in bp. ATP11: ATP11 protein family. GPCR: Serpentine type 7TM GPCR chemoreceptor. NAM: No apical meristem-associated C-terminal domain. TRX: Thioredoxin protein superfamily.

Table 2. Harbinger transposons studied in *Brassica* with TSDs, TIRs and positions in BAC sequences.

Name	Accession	Host	Size	TSDs	TIR (5′-3′)	Position
BoHARB1	AC240081.1	*B. oleracea*	3843	TAA	CAATAGGTCTGTTCGTTTGGTGCCCG CAGATTCCTGCGGCTG	5984-9826
BoHARB2	AC240081.1	*B. oleracea*	3755	TTA	GACCATCATTATCCC	53192-56946
BoHARB3	AC240089.1	*B. oleracea*	4063	AAG	GCTTAGAGCATGATTATC	86355-90417
BrHARB4	AC189588.2	*B. rapa*	3527	TAA	TTAATGGTTGCTTTA	34915-38440
BrHARB5	CU984545.1	*B. rapa*	2672	TAA	GAGCATCTTTATCCATG	36506-39177
BoN-HARB1	EU642504.1	*B. oleracea*	1199	TTA	GAGAATCTCCAAAAGAAACTCTAT	68290-69477
BrN-HARB2	AC189298.1	*B. rapa*	819	TAC	AATATGGTGAATTGAAATAGAAT	46497-47315
BoN-HARB3	AC240089.1	*B. oleracea*	514	TCA	ATTGTCAATCTCTAAGACCATCGTT	9672-10185

The *BoHARB3* was identified from *B. oleracea* accession 'AC240089.1' from 86355-90417 bp. This 4063 bp large *Harbinger* was found to be terminated by AAG TSDs and 18 bp imperfect TIRs (5′-GCTTAGAGCATGATTATC-3′) (Fig. 1; Table 2). The element showed A/T rich nuclotides (76%) in the terminal 400 nucleotides excluding TIRs at 3′ (C-terminal) end. The molecular structure of *BoHARB3* revealed that it encodes transposase protein in sub-terminal region of 3′ end. Besides a transposase two other proteins TRX and a GPCR family are encoded by it. This protein is located towards the C-terminal end of SANT protein domain and N-terminal end of transposase (TNP). BLASTN searches of *BoHARB3* transposase against *Brassicaceae* genome database in NCBI retrieved 140 copies with >90% homology in entire length. *BrHARB4* is autonomous *Harbinger* identified from the *B. rapa* accession 'AC189588.2'. The element ranged in size of 3527 bp with typical TAA TSDs and 15 bp TIRs (Table 2). A ~200 bp simple sequence repeat (CT_n) was detected 250 bp away from the start of 5′ TIR. The element was found to be A/T rich (60%) with several simple poly (AT) repeats. *BrHARB4* showed >50% homology in its entire length and >90% homology in transposase region of *BrHARB3*. The *BrHARB4* displayed a transposase domain in addition to SANT and NAM domains in its structure (Fig. 1). *BrHARB5* was isolated from *B. rapa* accession 'CU984545.1' from 36506-39177 bp flanked by TAA TSDs, 17 bp TIRs and high A/T content (60%) in its molecular structure. The *BrHARB5* displayed SANT and NAM associated protein domains, while lack a potential transposase (Fig. 1).

PCR amplification of *Harbinger* transposase in *Brassica*: The diversity and amplification pattern of *Harbinger* specific transposase was performed using 40 *Brassica* cultivars (Table 1). As the transposase in *BoHARB3* and *BoHARB4* elements shared >90% homology, a mutual primer pair BoHARB3/4F 5′-CGATGAGTACTTAAGAAGAC-3′ and BoHARB3/4R 5′-GGCAAGATTATGAGAGCATG-3′ was designed around the DDE motif to investigate *Harbinger* transposase diversity in *Brassica* genomes. Of the 40 *Brassica* accessions tested (Table 1), 566 bp *BoHARB3/BoHARB4* transposase was amplified from 38 diploids and polyploids *Brassica* (Fig. 2a). The only genomes failed to amplify the transposase were *B. rapa* accessions 'Pak Choy' and 'Vertus'. Very weak band was observed in *B. rapa pekinensis*, where PCR was repeated to gain strong bands at different annealing temperature. The amplified products in three *B. nigra* accessions suggest its presence in B-genome *Brassica*. In addition to the amplification of expected band, additional bands of ~350-550 bp were amplified from all six *B. oleracea* accessions. All the nine *B. juncea* (AABB), six *B. carinata* (AACC) accessions and four synthetic hexaploids amplified the *Harbinger* transposase (Fig. 2a).

Insertional polymorphism of *BrHARB5* in *Brassica* genomes: The defective (transposase deleted) *BrHARB5* was blasted against the GenBank database to collect other homologues but no significant hits were received from *Brassica* species except *B. rapa*. The question arises weather *BrHARB5* was unique to *B. rapa* or dispersed in other *Brassica* species? To answer the questions, the markers (primers) were developed; one pair from N-terminal end (BrHARB5F: 5′-CGCCATTGTTTCATGTGTGT-3′ and BrHARB5R: 5′-GCATTCAGATGATGTTGTGC-3′) and other from C-terminal end (BrHARB5F: 5′-GCACAACATCATCTGAATGC-3′ and BrHARB5R: 5′-GTACTACTGTCTACGTATGG-3′) of the insertion amplifying the 1516 bp N-terminal half (including 192 bp flanking region) and 1521 bp C-terminal half (including 153 bp flanking region) respectively (Fig. 2b-d). Both parts of *BrHARB5* were amplified in all the six *B. rapa* accessions. In contrast, no amplification was observed in any of *B. nigra* (BB genome) and *B. oleracea* (CC genome) accessions (Fig. 2b-c). This confirmed our hypothesis of A-genome specificity of *BrHARB5*. The amplification pattern of *BrHARB5* in allotetraploids and hexaploids further strengthens the A-genome specificity of the element, where only *B. juncea* (AABB), *B. napus* (AACC) and hexaploid *Brassica* (AABBCC) yielded the product, while *B. carinata* (BBCC) failed to amplify. Of the 9 *B. juncea*, 'Kai Choy' and 'Tsai Sim' accessions amplified the 1516 and 1521 bp N- and C-terminal parts (Fig. 2d) of *BrHARB5*. *B. napus* accessions except 'Last and Best' amplified both parts of *BrHARB5*. Similarly 2 hexaploids (AABBCC: *B. napus* x *B. nigra*) amplified the complete *BrHARB5*.

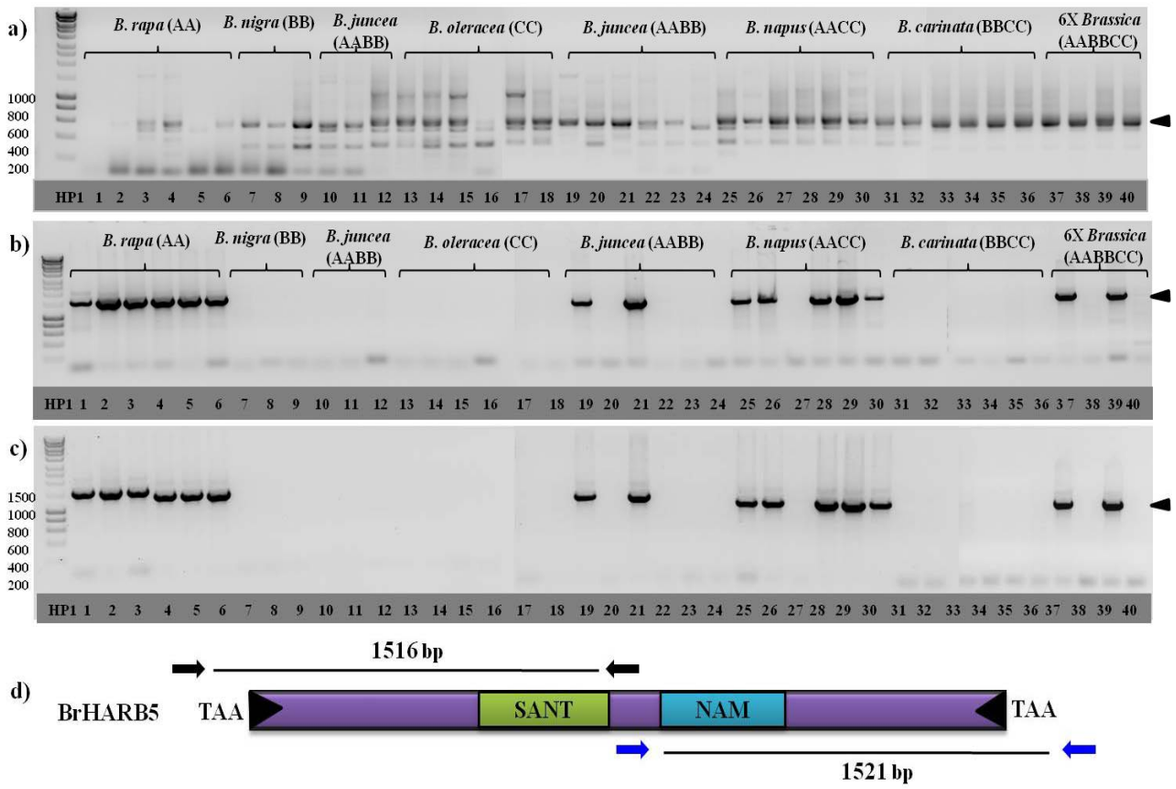


Fig. 2. PCR amplification of a) 566 bp *Brassica Harbingers* transposase. The transposase is present in most of *Brassica* genomes except accessions 1 and 5. b) N-terminal (first) half of *BrHARB5* (1566 bp) c) C-terminal (last) half of *BrHARB5* amplified from A-genome and its allotetraploids (AABB, AACC) and hexaploids (AABBCC). Arrow heads are indicating the expected product sizes. The numbers below are the identifiers of the *Brassica* accessions listed in Table 1d) Showing the position of markers (primers) with product sizes from *BrHARB5*.

Structural features of non-autonomous *Harbingers* in *Brassica*: Three short non-autonomous (<1.2 kb) elements were identified from *Brassica* genomes. The first element was detected from *B. oleracea* accession 'EU642504.1' as an insertion from 68290-69477 bp within the BAC sequence. The element *BoN-HARB1* (1199 bp) is flanked by 3 bp TSDs (TTA) and 24 bp perfect TIRs (Table 2). The element is highly AT rich (76%) with dispersed poly AT sequences. It captures a ~500 bp NADH dehydrogenase subunit (ND5) domain. Using *BoN-HARB1* as a query in GenBank database, 365 sequences showed homology to the element with half elements showing >75% identity in their entire lengths. The members of this family range in sizes from 1042-1215 bp, terminated by TAA/TTA TSDs and 24-25 bp TIRs (Table 3), which are highly conserved with the exception of 1-3 bp mismatch, otherwise the 24 bp TIRs are similar in their entire lengths in all copies (Fig. 3). Another non-autonomous *Harbinger BrN-HARB2* was isolated from *B. rapa* accession 'AC189298.2' from 46497-47315 bp. The 819 bp element is terminated by TAC TSDs and 23 bp TIRs. *BoN-HARB3* was detected as an insertion in *B. oleracea* BAC clone (AC240089.1) from nucleotide position 9672-10185. The element is 514 bp displaying 3 bp TSDs and 26 bp imperfect TIRs (Table 2).

Insertion polymorphism of non-autonomous *Harbingers* in *Brassica*: The insertion polymorphisms of *Brassica* non-autonomous *Harbingers* were performed by using 'Sequence Specific Amplification Polymorphism' (SSAP) markers designed from flanking regions of insertions. The higher and lower bands were achieved on the basis of presence or absence of insertions at specific loci. The BoNHARB1F (5′-ACTAGCCATTTCCATCTTCT-3′) and BoNHARB1R (3′-GTATTCACTTGTAGTGTTTG-5′) primer pair was used to amplify 1199 bp *BoN-HARB1* element with a product size of ~1357 bp including the flanking regions (Fig. 4a). The amplification of *BoN-HARB1* was not observed in any of A-genome, but B and C-genome *Brassica* diploids yielded the expected bands. All the three *B. nigra* (B-genome) and six *B. oleracea* (C-genome) accessions amplified the ~1357 bp segments. Similarly, four *B. napus* (Mar, Last and Best, Fortune, Drakker) and six *B. carinata* accessions amplified the *BrN-HARB1* elements. Another primer pair BoNHARB2F (5′-ACATGCATAGATTGCGCTTG-3′) and BoNHARB2R (3′-TTTTCACATTCGGCATGAGT-5′) was designed to amplify a 819 bp *BoN-HARB2* element with a product size of 1100 bp including ~180 bp flanking regions (Fig. 4b). The primer amplified the desired bands (weak) from two *B. rapa* (Pak Choy, Chinese Wong Bok) and four *B. juncea* (NARC-I, NATCO, W3, Varuna) accessions. All the other accessions failed to amplify 819 bp *BoN-HARB2* indicating its absence.

Table 3. List of non-autonomous *BoN-HARB1* and its homologues in *Brassica* with TSDs and TIRs.

Name	Accession	Species	Size	TSDs	TIR (5′-3′)
BoN-HARB1-1	EU642504.1	*B. oleracea*	1199	TTA	GAGAATCTCCAAAAGAAACTCTAT
BoN-HARB1-2	AC183494.1	*B. oleracea*	1095	TTA	TAGCATCTCCAAAAGACACTCTAT
BoN-HARB1-3	AC183493.1	*B. oleracea*	1096	TTA	GAGCATCTCCAAAAGACACTCTAT
BrN-HARB1-4	AC189475.2	*B. rapa*	1212	TAA	GAGCATCTCCAAAAGAAACTCTAT
BrN-HARB1-5	AC189364.2	*B. rapa*	1212	TCA	GAGCATTTCCAAAAGAAACTCTAT
BrN-HARB1-6	AC189237.1	*B. rapa*	1215	TTA	GAGCATCTCCAAAAGAAACTCTAT
BrN-HARB1-7	AC189430.2	*B. rapa*	1213	TAA	GAGCATCTCCAAAAGAAACTCTAT
BrN-HARB1-8	AC232512.1	*B. rapa*	1136	TTA	CAGCATCTCCAAAAGAAACTCTAT
BrN-HARB1-9	AC189375.2	*B. rapa*	1102	TTA	GAGCATCTCCAAAAAATATTCTAT
BrN-HARB1-10	AC189300.2	*B. rapa*	1086	TAA	GAGCATCTCCAAAAGACACTCTAT
BrN-HARB1-11	AC189225.2	*B. rapa*	1063	TAA	GAGTATCTCCAAAAGACACTCTAT
BrN-HARB1-12	AC232514.1	*B. rapa*	1208	TAA	GAGCATCTCCAAAAGAAACTCTAT
BrN-HARB1-13	AC189183.2	*B. rapa*	1117	TAA	GAGCATCTCCAAAAGAAACTCTAT
BrN-HARB1-14	AC189592.2	*B. rapa*	1042	TAA	GAACATCTCCAAAAGAAACTTTAT
BnN-HARB1-15	AC236787.1	*B. napus*	1095	TTA	TAGCATCTCCAAAAGACACTCTAT

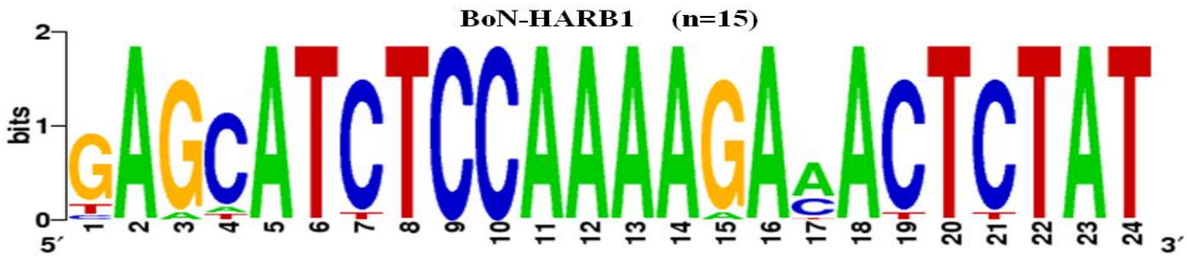


Fig. 3. WebLogo representing TIRs of *BoN-HARB1* elements generated from TIRs of 15 elements. Nucleotides 1, 3, 4 and 17 are most variable, while others particularly 8 to 14 are highly conserved among various elements.

The phylogenetic relationship of *Brassica* and related plant *Harbingers*: The phylogenetic analysis of 28 DDE catalytic transposases of *Brassica* and related plants revealed that they could be divided into dicot and monocot clades, further resolving them into five sub-clades (Fig. 5). The clades were monophyletic with minor heterogeneity in their transposase regions. Clade 1 designated as *HARB1_T. aestivum* clustered the 7 monocot transposases including two previously known *HARBI1* from *Triticum aestivum* and *HARBI2* from *Zea mays*. The transposase of *T. aestivum*, *Aegilops tauschii* and *Brachypodium distachyon* constituted one family, while *Z. mays* clustered in other family. The second clade grouped 21 dicot plant transposases, further dividing them into four sub-clades or families. The transposase of *Brassica Harbingers* have shown high homology by clustering in single '*BoHARB3-Brassica*' family. The transposase from *BoHARB3* and *BrHARB4* from *B. oleracea* and *B. rapa* respectively clustered on sister branch, while all other *Brassica* transposases clustered on nearby branches constituting the single family (Fig. 5). The 5 *Arabidopsis thaliana Harbingers* retrieved from NCBI and 1 known *Harbinger* (*ATIS112A*) from Repbase database clustered together in second sub-clade named '*ATIS112A-A. thaliana'*. The 2 *AtHARB* clustered on one branch, while other 4 clustered in other group. The third sub-clade named '*MTISI12A-M. truncatula'* clustered elements from *Medicago truncatula* and *Camelina sativa*. All the three transposase from *Malus domestica* grouped in fourth sub-clade (*HARBI1-M. domestica*). The high homologies observed in monocot and dicot *Harbinger* transposase and genus-specific groups within *Brassica* and *Arabidopsis* suggested their common ancestry.

Comparative genomics of *Brassica* and related plant transposases: The alignment of *Harbinger* transposase around DDE domain (~200 aa) revealed that the transposase from various plants showed high homology and conserved regions; maximum homology was observed among *Brassicaceae* members (Fig. 6b). The comparison of transposase from monocot and dicot *Harbinger* elements showed homology in their entire lengths with few varied regions. The conserved $D_{88}D_{35}E$ triad was detected from all transposase sequences (Fig. 6a, b). Beside the DDE conserved triad, few other conserved Aspartic (D) and Glutamic acid (E) residues were observed at variable intervals in aligned sequences. In general the structure of conserved residues was $D_{20}\mathbf{D}_{28}E_{21}D_5D_{37}\mathbf{D}_{35}\mathbf{E}$, where the bold letters are indicating the DDE triad and the numbers are indicating the amino residue spacing. The most conserved transposase motifs were 'GSI/LDCMHW' (aa position 35-42), 'LEAVA' (aa position 67-71), 'WIWH' (aa position 76-79), 'YYLT/ADGIYP' (aa position 124-132) and 'RKDV/IERAFG' (aa position 160-168) etc (Fig. 6a, b). Beside these, several other 1-3 aa conserved motifs were also found. Five highly conserved blocks were identified from the aligned sequences showing the most conserved regions (Fig. 6a,b). Several amino acid polymorphisms were also observed, with frame shift mutations and stop codons in few such as *M. domestica* sequences. The alignment confirmed a lot of homology within *Brassica* and related plant transposases, with varied regions separating them into their respective clades (monocot and dicot) and families in evolutionary analysis.

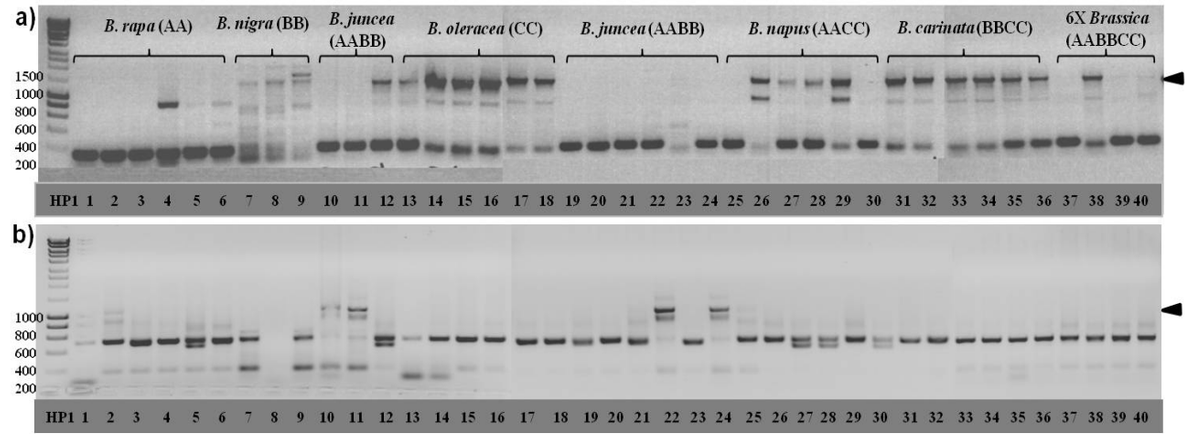


Fig. 4. Insertional polymorphism of non-autonomous *Harbingers*. a) Upper bands (1357 bp) amplifying 1199 bp *BoN-HARB1* b) upper bands (1100 bp) amplifying 819 bp *BrN-HARB2* from various *Brassica* lines. Arrow heads indicating the products, numbers below are the identifiers of the *Brassica* accessions listed in Table 1.

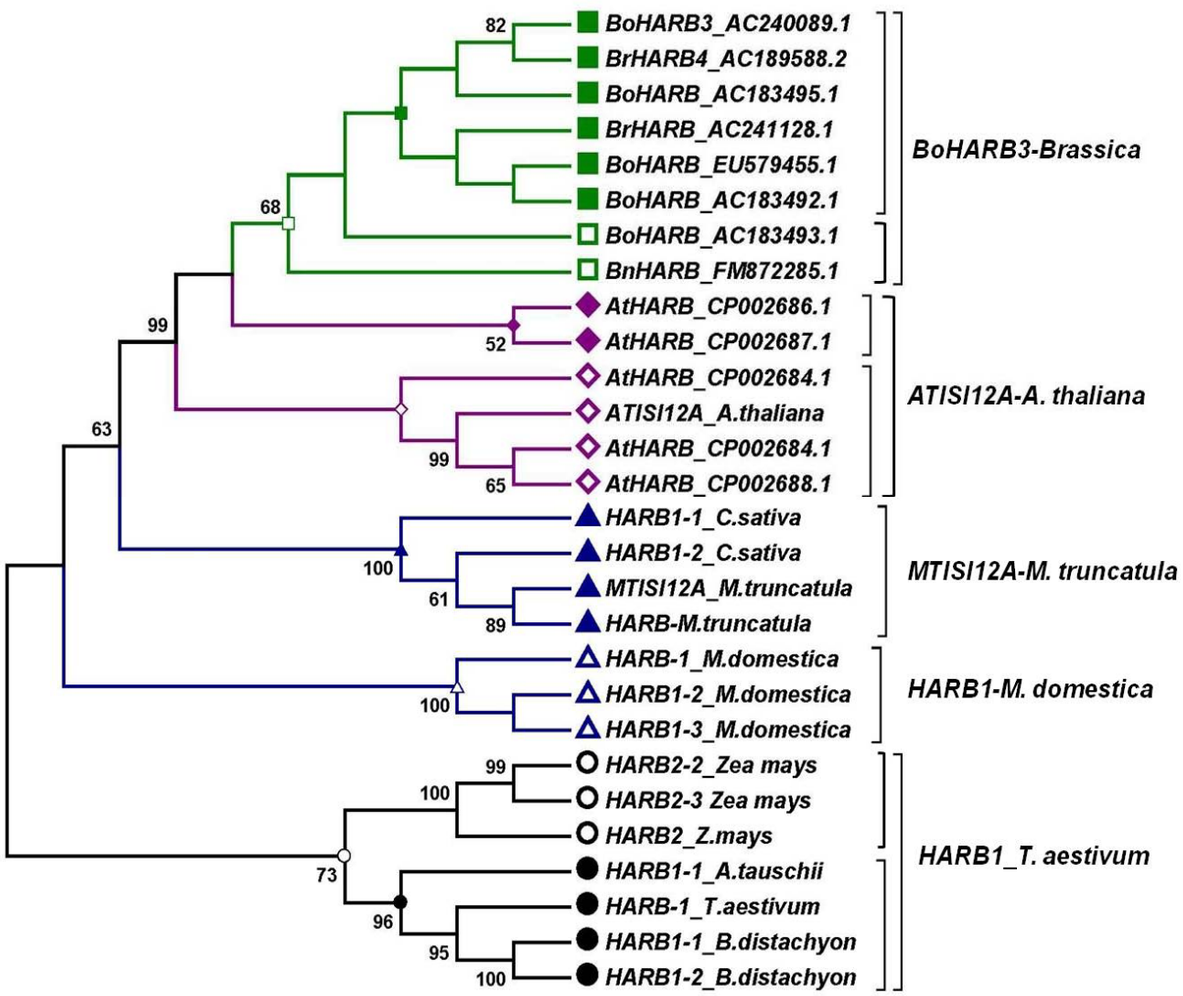


Fig. 5. Phylogenetic Neighbour-Joining tree of *Brassica* and related *Harbinger* transposase with 1000 bootstrap values (shown as %). The names show BAC accessions or for non-*Brassicaceae* elements, species names; family names (right) on the basis of the well known *Harbinger*. *Arabidopsis* (*At*) and *Brassica* (*Br/Bo/Bn*) sequences are named according to the genera and species initials followed by the GenBank accession number. Various clades, sub-clades, and families are represented with open and filled shapes. *A. thaliana*: *Arabidopsis thaliana C. sativa*: *Camelina sativa. M. truncatula*: *Medicago truncatula. M. domestica*: *Malus domestica. A. tauschii*: *Aegilops tauschii. T. aestivum*: *Triticum aestivum. B. distachyon*: *Brachypodium distachyon.*

Table 4. Size and protein domain organizations of *Brassica* and related *Harbingers*. The Known *Harbingers* were collected from Repbase database.

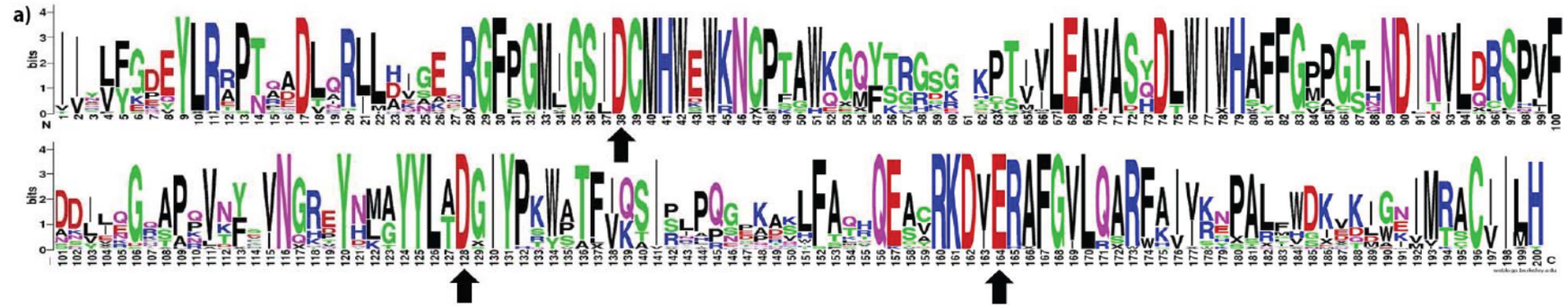
No.	Element Name	Plant Species	Size	Domains (5′-3′)	Reference
1.	*BoHARB1*	*Brassica oleracea*	3837	SANT	Present Study
2.	*BoHARB2*	*Brassica oleracea*	3749	TRX-ATP11	″
3.	*BoHARB3*	*Brassica oleracea*	4057	SANT-GPCR-TNP	″
4.	*BrHARB4*	*Brassica rapa*	3521	SANT-NAM-TNP	″
5.	*BrHARB5*	*Brassica rapa*	2672	SANT-NAM	″
6.	*HARBINGER*	*Arabidopsis thaliana*	5382	TNP	Repbase database (Jurka *et al*., 2005)
7.	*ATIS112A*	*Arabidopsis thaliana*	5099	TNP	″
8.	*HARB-3_Stu*	*Solanum tuberosum*	4212	SANT-TNP	″
9.	*Harbinger-1_VV*	*Vitis vinifera*	4378	SANT-TNP	″
10.	*MTISI12A*	*Medicago truncatula*	3914	SANT-TNP	″
11.	*HARB-1_Mad*	*Malus domestica*	2818	TNP	″
12.	*HARB-2_ZM*	*Zea mays*	6231	TNP-NAM	″
13.	*HARB-1_TA*	*Triticum aestivum*	2161	TNP	″
14.	*HARB-1_OS*	*Oryza sativa*	5166	SANT-NAM-TNP	″
15.	*HARB-10_SBi*	*Sorghum bicolour*	5934	TNP-SANT-CVV	″

Discussion

Mobile DNA elements or TEs although constituting major proportions of most prokaryotic and eukaryotic genomes are not fully characterized in several genomes. Among DNA transposons, *Harbinger* elements are not investigated in detail due to lack of knowledge or very short or ill defined characteristics hallmarks (TSDs, TIRs). In recent years they were identified from few plant, animal and fungal genomes (Kapitanov & Jurka, 2004; Pereira *et al*., 2013), but less data is available for them. In present study, we identified 8 *Harbinger* transposons, of which only 2 are complete with putative transposase, 3 lack transposase but displayed other domains (Fig. 1) and rest 3 are non-autonomous (lack domains) elements (Table 2). The elements ranged in sizes from 0.5-4 kb with 3 bp TSDs and 15-42 bp TIRs (Table 2). Several other *Harbingers* identified from other plants showed similar sizes as *PIF/Harbingers* identified from *M. trancatula* ranged from 3.1 to 6 kb (Grzebelus *et al*., 2007). *BoHARB5* (2.6 kb) showed similar size to a 2.5 kb *DcMaster1* from *Daucus carota,* which was found inserted in the first intron of carrot vacuolar acid invertase isozyme-II gene. The insertion was characterized by TTA TSDs, 22 bp TIRs, 43 bp imperfect sub-terminal regions and lack transposase (Grzebelus *et al.*, 2006), as observed in *BoHARB1, BoHARB2* and *BrHARB5* elements in present study. Another element *Boto* from *Moniliophthora perniciosa* was identified having a size of 3080 bp flanked by 45 bp TIRs (Pereira *et al*., 2013). The present study involved the identification of non-autonomous *Harbingers BoN-HARB1, BrN-HARB2* and *BoN-HARB3* with a size of 1199, 819, 514 respectively (Table 2). Several short non-autonomous *Harbingers* were identified from *Medicago* genome with similar sizes (Grzebelus *et al*., 2007; Markova, 2014). Their distribution and proliferation in *Brassica* genome suggested that their autonomous partners are assisting in their proliferation and mobilization by providing their transposase enzymatic machinery.

The present and previous studies confirmed the 3 bp TSDs of all *Harbingers,* although few TSDs were different from putative *Harbinger* TAA TSDs. The TIRs of the *Brassica Harbingers* ranged in size from 10 to 42 bp (Table 2). The size of *Harbinger* TIRs investigated in other organisms varies and ranged from 10 to 45 bp as observed in *Boto* (Pereira *et al*., 2013) and rice *PIF*-like elements (Zhang *et al*., 2004). In *Medicago*, the TIRs of all the six *Harbingers* families ranged from 14 to 22 bp (Grzebelus *et al*., 2007). Although the TIRs could be larger than TIRs identified here as Kapitanov and Jurka (2004) investigated TIRs of various *Harbingers* from 10-700 bp. In all previous studies, the TIRs mostly start with GGG or GNG (where N is any other nucleotide), but in *Brassica Harbingers*, no such correlation exists. The *Harbingers* showed less activity and abundance in *Brassica* genome as compared to other DNA transposons in *Brassica*. There might be two possibilities for their less abundance i) their lost from *Brassica* genome during evolutionary timeframe ii) horizontal transfer limited their transposition rates in *Brassica* genome. The number of copies investigated for other DNA transposon superfamilies like *CACTA*, *hATs*, *Mariner* and *Mutator* were high in *Brassica* genome (Nouroz, 2012, Nouroz *et al*., 2015c).

The amplification of *Harbinger* transposase showed that *Harbinger* transposons are ancient superfamily of DNA transposons and were present in A, B and C-genome *Brassicas* before their divergence from a common ancestor (Fig. 2a). The lack of amplification in *B. rapa* accessions (Pak Choy and Vertus) might be due the difference in annealing temperatures or there might be a possibility of mutation at primer sites. Similar amplification pattern of *Boto* element was identified in various genome isolates of *M. perniciosa* (Pereira *et al*., 2013). The *Harbinger* based site specific insertion polymorphisms were observed, where the insertion was amplified in few accessions, while lack in others providing best molecular markers for *Brassica* accession identification (Fig. 2b,c; 4a,b). The IRAP, REMAP (Kalender & Schulman, 2006), RAPID (Wu *et al*., 2014), DNA transposons *CACTA* (Alix *et al*., 2008), *Brassica* Microsatellites (Sadia *et al*., 2010), *MITEs* (Yakoov *et al*., 2012; Nouroz *et al*., 2015b) and present *Harbingers* related transposon based molecular markers are highly informative for varieties/accessions identification or to investigate the biodiversity and evolution of genomes.



b)

```
                    10        D 20        30        D 40        50        60        E 70  D       80        D90       100
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
BoHARB3_AC240089.1  IIYLFGDEYL RRPTRADLQRLIDVGEYRGFPGMIGSIDCMHWEWKNCPTAWKGQYSRGSG-KPTIVLEAVASYDLWIWHAFFGPPGTLNDINVLDRSPVF
BrHARB4_AC189588.2  IIYLFGNEYLRRPTPADLQRLLDVGEYRGFPGMIGSIDCMHWEWKNCPTAWKGQYSRGSG-KPTIVLEAVASYDLWIWHAFFGPPGTLNDINVLDRSPVF
BoHARB_AC183493.1   IISLFGEEYLRRPTPADLQRLLDIGEYRGFPGMVGSIDCMHWEWKNCPTAWKGQYTRGLG-KPTIVLEAVASYDLWIWHAFFGPPGTLNDINVLDRSPVF
BoHARB_EU579455.1   IINLFGDEYLRRPTPADLQRLLDISKQRGFSGMIGSIDCMHWEWKNCLTAWKGQYSRGSG-KPTIVLEAVASYDLWIWHAFFGPPCTLNDINVLDRSPVF
BrHARB_AC241128.1   IIYLFSDEYLRRPTQADLQRLLDIADCRGFPGMIGSIDCMHWEWKNCPTAWKGQYSRGSG-KPTIVLEAVASYDLWIWHSFFGPPGTLNDINVLDRSPLF
BoHARB_AC183495.1   IIYLFGDEYLRRPTLADLQRLLDIGEHRGFSGMIGSIDCMHWKWKNCPTAWKCQYSHGSG-KPTIXXEAVASYDLWIWHAFFGPPGTLNDINVLDRSPVF
BoHARB_AC183492.1   IIYLFGDEYLRRPTLADLQRLLDIGKERGFSGMIGSIDCMHWEWKNCPTTWKGMYSRGTG-KPTIVLEAVASYDLWIWHAFFGPPGTMNDLNILDRLPVF
BnHARB_FM872285.1   IIYLFGDEYLRRPTTADLQILLDIGEARGFPGMIGSIDCMHWEWKNCPTAWKGQFTRGSG-KPTIVLEAVASYDLWIWHAFFGPPGTLNDINVLDRSPVF
AtHARB_CP002684.1   IINLFGDEYLRRPTRDDLIRLLHIGEQRGFPGMIGSIDCMHWEWKNCPTAWKGQYTRGSG-KPTIVLEAVASQDLWIWHAFFGPPCTLNDINVLDRSPVF
AtHARB_CP002684.1   IIYLFGKEYLRRPTRQDLKRLLRIGELRGFLGMTGSIDCMHWEWKNCPTAWKEQYTRGSG-KPTIVLEAVASQDLWIWHICFGPLGTLNDINIIDLSLVF
AtHARB_CP002686.1   IISFFGDEYLRAPTATDLRRLLNIGKIRGFPGMIGSLDCMHWEWKNCPTAWKGQYTRGSE-XXTIVLEAIASQDLWIWHVFFGPPGTLNDINILDRSPIF
AtHARB_CP002687.1   IISLFGDEYLRTPTQADLQRLLDIREIRGFSGMIGSIDCMHWEWKNCPTSWKGQYTRGSG-KPTIVLEAVASQDLWIWHAFFGPLGTLNDINILDRSPVF
AtHARB_CP002688.1   TINLFGDEYLRRPTRDDLIRLLHIGEQRGFPGMIGSIDCMHWEWKNCPTAWKGQYTRGSG-KPTIVLEAVASQDLWIWHAFFGPPGTLNDINVFDRSPVF
ATISI12A_A.thaliana IINLFGDEYLRRPTRDDLIRLLHIGEQRGFPGMIGSIDCMHWEWKNCPTAWKGQYTRGSG-KPTIVLEAVASQDLWIWHAFFGPPGTLNDINVLDRSPVF
HARB1-1_C.sativa    IIRLYKQEYLRAPTQDDLQRILHVSEMRGFPGIIGSIDCMHWEWKNCPTAWVGQFTRGDKGTTTVILEAVASHDLWIWHAFFGCPGTLNDINVLDRSPVF
HARB1-2_C.sativa    IIQLYEKEYLRAPTQEDLQRILHVSEMRGFPGMIGSIDCMHWEWKNCPTAWKGQFTRGDKGTTTVILEAVASHDLWIWHAFFGCPGTLNDINVLDRSPVF
MTISI12A_M.truncatula IIRLYEEVYLRAPTQDDLQRILHVSEMRGFPGMIGSIDCMHWEWKNCPKAWEGQFTRGDR-EPTVILEAVASHDLWIWHAFFGCPGTLNDINVLDRSPVF
HARB-M.truncatula   IIRLYEQVYLRAPTQDDLQRILHVSEMRGFLGMIGSIDCMHWEWKNCPKAWEGQFTRGDKGTTTVILEADASHDLWIWHAFFGCPGTLNDINVLDRSPVF
HARB-1-2_M.domestica IVXVYKDEYXXEPNZXDLNR-LHKAEDXGFPGMXGXLDCMHWDWKNXXTGWQXXFSXRSR-KPTXVLEAVASYDTWIXHAFFGVXGSQNDITVLXXSPXF
HARB-1_M.domestica  IVQVYKDEYLREPNQEDLNRLLHKAEDRGFPGMIGSLDCMHWDWKNCPTGWQGGFSGRSR-KPTVVLEAVASYDTWIWHAFFGVPGSQNDITVLGRSPLF
HARB1-3_M.domestica IVQVYKDEYLREPNQEDLNRLLHKAEDXGFPGMIGSLDCMHWBWKNCPTGWQXXFSGRSR-KPTXVLEAVASYDTWIXHAFFGVPXSQNDITVLXRSXLF
HARB2-2_Zea mays    IIDVYGPYYLRAPNAADVARLLAEGEQRGFPGMLGSIDCMHWEWRNCPSAWKGMFTGRGK-HPSMILEAVASHDLWIWHAYFGMPGSNNDINVLQRSPIF
HARB2-3 Zea mays    IIDVYGPYYLRAPNAADVARLLAEGEQRGFPGMLGSIDCMHWEWRNCPSAWKGMFTGRGK-HPSMILEAVASHDLWIWHAYFGMPGSNNDINVLQRSPVF
HARB2_Z.mays        IIDVYGPYYLRAPNAADVARLLAEGEQRGFPGMLGSIDCMHWEWRNCPSAWKGMFTGRGK-HPSMILEAVASHDLWIWHXFFGMPGSNNDINVLQRSPVF
HARB1-1_A.tauschii  VIEVFGPEYLRQSTTADTERLLATNAARGFPGMLGSIDCMHWEWKNCPFAWQSQYKGHVN-GCTVILEAVASLDLWIWHSFFSMAGSHNDIDVLQRFPVF
HARB-1_T.aestivum   VVAVFGPQYLRSPNAEDLLG-LAQNAARGFPGMLGSIDCMHWKWKNXPFAWQGMYKGAKG-GXSVVLEAVATQDLWIWHSFFGMPGTHNDINVLQCSPVF
HARB1-1_B.distachyon IVAVFGEQYLRTPTAEDTARIMAQNVERGFPGMLGSIDCMHWSWKNCPFAHQGMYKSHKG-SCSVMLEAVADQDLWIWHAFFGMAGSHNDINVLQCSHVF
HARB1-2_B.distachyon IVAVFGEQYLRTPTAKDTARIMAQNAERGFPSMLGSIDCMHWSWKNCPFAHQGMYKSHKE-SCGVVLEAVADQDLWIWHAFFGMAGSHNDINVLQCSNVF
```

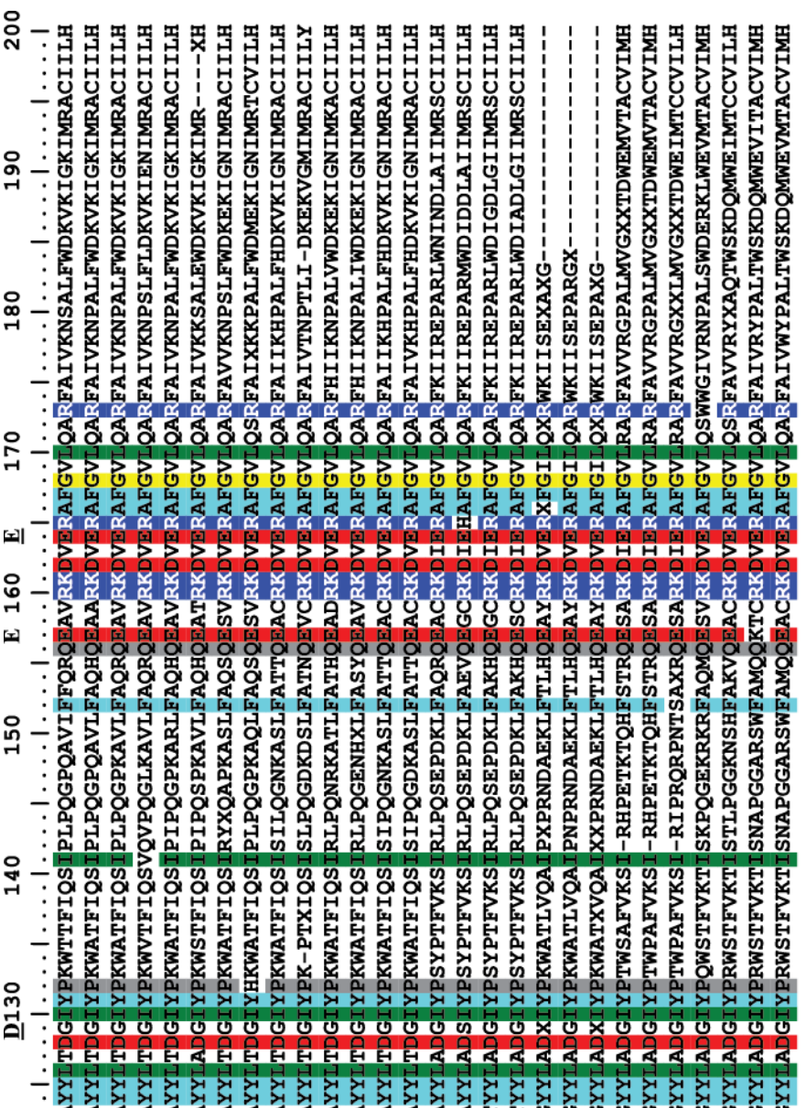


Fig. 6. a) WebLogo of 200 aa transposase indicating DDE motif with arrows. The letters with upper heights showed conserved, while short heighted letters are representing the variable regions. *A. thaliana*: *Arabidopsis thaliana C. sativa*: *Camelina sativa*. *M. truncatula*: *Medicago truncatula*: *M. domestica*: *Malus domestica*. *A. tauschii*: *Aegilops tauschii*. *T. aestivum*: *Triticum aestivum*. *B. distachyon*: *Brachypodium distachyon*. b) Multiple sequence alignment of transposase encoded by *Brassica* and related *Harbinger* elements. The conserved $D_{88}D_{35}E$ triad is indicated by the underlined letters at top. The most conserved regions are underlined indicating five conserved blocks in transposase. Small insertions were deleted without altering the frame. Asterisks show the stop codons, letter X indicates an incomplete codon and dashed lines represent gaps or ends of incomplete sequences.

Brassica and other plant *Harbingers* have showed two ORFs with diversity in their proteins domain organization (Table 4). The present study confirmed the previous studies demonstrating that all major *Harbingers* from eukaryotic genomes encode two proteins but few additional domains can also be detected (Kapitonov & Jurka, 2004; Markova, 2014). *BoHARB1* only encodes a SANT protein, while *BoHARB2* captures thioredoxin (TRX) and ATP11 protein domains only (Table 4). *BoHARB3* and *BrHARB4* encode transposase and SANT protein domains with one additional protein GPCR and NAM respectively, while *BrHARB5* only encode SANT and NAM domains. The domain organization of *Harbinger* elements from other species revealed similar range of variation in number and nature of ORFs (Table 4). Examples include the 5.3 kb *HARBINGER,* 5.0 kb *ATIS112* element from *A. thaliana* and 2.8 kb *HARB-1_Mad* from *M. domestica* that only encode a transposase in their molecules. The *HARB-3_Stu* from *Solanum tuberosum*, Harbinger-*1_VV* from *Vitis vinifera* and *MTISI12A* from *M. truncatula* encode transposase and SANT protein domains, which are present in majority of plant *Harbingers*. A 6.2 kb *HARB-2_ZM* from *Z. mays* encodes a transposase and NAM family of proteins. A 2.1 kb large *T. aestivum* element *HARB-1_TA* only encodes a transposase, while *HARB-1_OS* from *Oryza sativa* and *HARB-10_SBi* from *Sorghum bicolor* encode SANT and transposase domains (Table 4).

The phylogenetic analysis of *Brassica* DDE transposase and related sequences clustered them into monocot and dicot clades further resolving them into 5 sub-clades or families (Fig. 5). *Brassica*, *Arabidopsis* and *Malus* constituted species specific groups, while *Z. mays*, *Triticum*, *Aegelops* and *Brachypodium* transposase grouped together. The evolutionary relationship of prokaryotic and eukaryotic transposase from various organism revealed that the plant, animal, fungal and bacterial transposases despite having homologies clustered in their respective groups (Kapitanov & Jurka, 2004). In present study, the known *ATISI2A* element grouped with *Arabidopsis Harbingers* constituting sister family with *Brassica Harbingers.* In previous evolutionary studies, *ATISI2A* formed sister branch with *Pong*-like element (Kapitonov & Jurka, 2004), thus revealing the relationship of *Pong* and *Arabidopsis/Brassica Harbingers*.

The multiple sequence alignment of *Brassica* and related transposase sequences of the present manuscript revealed identification of five conserved blocks (Fig. 6b). The comparative genomics of transposase domains collected from various plants, animal, fungal and bacterial genomes also revealed the presence of such conserved blocks, considered to be catalytic hotspots for nuclease/ligase reactions necessary for transposition (Kapitanov & Jurka, 2004). The analysis revealed that *Brassica* and related plants showed homology to *Pong* and *ATISI12A* sequences. The $D_{88}D_{35}E$ motif was detected from almost all sequences (Fig. 6a,b), which showed similarity to the other *Harbinger* or *Pong*-like elements with $DD_{35}E$ motif (Kapitanov & Jurka, 1999, 2004; Zhang *et al*., 2004). The DDE spacing in *Brassica* transposase was different from the spacing detected in several other plant *PIF/Harbinger* elements, i.e. $DD_{47}E$ or $DD_{48}E$ as investigated in *Boto* element (Pereira *et al*., 2013).

Conclusion

The present study is about detail characterization and diversity of novel *Brassica Harbingers*, a less abundant, ancient, but evolutionary active transposon superfamily. Our detailed characterization in *Brassica* showed the diversity in structure of *Harbinger* i.e. TSD sequence, TIR sizes, ORF composition and DDE transposase, which are characteristic of TE superfamilies and parallel the structures found in other well-analysed groups such as the *Triticeae* and *Brassicaceae.* The genome specificities of some of the *Harbinger* elements suggest that they will be valuable as probes for *in situ* hybridization to identify chromosome introgression and recombination events in hybrids (like the C-genome CACTA of Alix *et al*., 2008), but with the prospect of greater specificity and to the genomes. Since the PCR amplifications from different accessions within single species are sometimes showing polymorphisms, there is the potential to exploit these robust PCR markers for varietal identification, and perhaps for transposon-tagging of genes in appropriate populations as in systems based on *En/Spm* and *Ac/Ds* elements.

Acknowledgements

The authors are thankful to Hazara University Mansehra and Higher Education Commission, Pakistan for funding to work in highly reputed Molecular Cytogenetics laboratory, Department of Biology, University of Leicester, UK. We also pay our gratitude to Warwick Research Centre, UK and Dr. Xian Hong Ge, University of Wuhan, China for providing seeds/DNA material.

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.

Altschul, S.F., T.L. Madden, A.A. Schaffer. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*., 25: 3389-3402.

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

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

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

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.

Grzebelus, D and P.W. Simon. 2009. Diversity of *DcMaster*-like elements of the *PIF/Harbinger* superfamily in the carrot genome. *Genetica,* 135: 347-353.

Grzebelus, D, Y.Y. Yau and P.W. Simon. 2006. Master: a novel family of *PIF/Harbinger*-like transposable elements identified in carrot (*Daucus carota* L.). *Mol. Genet. Genomics,* 275: 450-459.

Grzebelus, D., S. Lasota, T. Gambin, G. Kucherov and A Gambin. 2007. Diversity and structure of *PIF/Harbinger*-like elements in the genome of *Medicago truncatula*. *BMC Genomics,* 8: 409.

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.

Hancock, C.N., F. Zhang and S.R. Wessler. 2010. Transposition of the *Tourist*-MITE *mPing* in yeast: an assey that retains key features of catalysis by the class II PIF/Harbinger superfamily. *Mob. DNA*, 1: 5.

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

Jiang, N., Z. Bao, X. Zhang, H. Hirochika, S.R. Eddy, S.R. McCouch and S.R. Wessler. 2003. An active DNA transposon family in Rice. *Nature*, 421: 163-167.

Jurka, J. and V.V. Kapitonov. 2001. PIFs meet Tourists and Harbingers: a superfamily reunion. *Proc. Natl. Acad. Sci. USA,* 98: 12315-12316.

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: 462-7.

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. 1999. Molecular paleontology of transposable elements from *Arabidopsis thaliana*. *Genetica,* 107: 27-37.

Kapitonov, V.V. and J. Jurka. 2004. *Harbinger* transposons and an ancient HARBI1 gene derived from a transposase. *DNA Cell Biol.,* 23: 311-324.

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.

Markova, D.M. 2014. Evolutionary Dynamics of *PIF* and *Pong*-like Transposable Elements in Triticeae. PhD Thesis. Department of Biological Sciences, University of Illinois, Chicago.

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, Department of Biology, University of Leicester, UK.

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

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

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.

Pereira, J.F., A.P.M.M. Almeida, J. Cota, J.A. Pamphile, G.F. da Silva, E.F. de Araujo, K.P. Gramacho, S.H. Brommonschenkel, G.A.G. Pereira and M.V. de Queiroz. 2013. *Boto*, a class II transposon in *Moniliophthora perniciosa*, is the first representative of *PIF/Harbinger* superfamily in phytopathogenic fungus. *Microbiol*., 159: 112-125.

Sadia, M., M.A. Rabbani, M.S. Masood, S.R. Pearce and S.A. Malik. 2010. Inter species testing of *Brassica* microsatellites available in public domain and their potential utilization for comparative genomics in cruciferae. *Pak. J. Bot*., 42(6): 3875-3885

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.

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. 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.,* 8(12): 973-982.

Wu, J., J. Wang, W. Du, Z. Wang, L. Chen, C. Liu, J. Zhao, Q. Yang and X. Chen. 2014. Production of a PCR-based marker for detecting Psathyrostachys Huashanica Keng chromosomes in a wheat background. *Pak. J. Bot*., 46(6): 2083-2088.

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

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

Zhang, X., C. Feschotte, Q. Zhang, N. Jiang, W.B. Eggleston and S.R Wessler. 2001. P instability factor: an active maize transposon system associated with the amplification of Tourist-like MITEs and a new superfamily of transposases. *Proc. Natl. Acad. Sci., USA,* 98: 12572-12577.

Zhou, M.B., J.J. Lu, H. Zhong, X.M. Liu and D.Q. Tang. 2010. Distribution and diversity of *PIF*-like transposable elements in the Bambusoideae subfamily. *Plant sci*., 179: 257-266.

Zhou, M.B., X.M. Liu and D.Q. Tang. 2012. PpPIF-1: first isolated full length PIF-like element from the bamboo *Phyllostachys pubescens*. *Genet Mol Res*., 11: 810-820.

(Received for publication 1 March 2015)