Plant Breed. Biotech. 2016 (August) 4(3):285~296
http://dx.doi.org/10.9787/PBB.2016.4.3.285
Online ISSN: 2287-9366
Print ISSN: 2287-9358

RESEARCH ARTICLE

Next-Generation Sequencing Based Transposon Display to Detect High-Throughput Insertion Polymorphism Markers in *Brassica*

Sampath Perumal[1], Nomar Espinosa Waminal[1,2], Jonghoon Lee[3], Nur Kholilatul Izzah[1,4], Mina Jin[5], Beom-Soon Choi[6], Tae-Jin Yang[1,7]*

[1]Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea
[2]Department of Life Science, Chromosome Research Institute, Sahmyook University, Seoul 01795, Korea
[3]Joeun Seed, Goesan 28051, Korea
[4]Indonesian Industrial and Beverage Crops Research Institute, Sukabumi 43357, Indonesia
[5]Genomics Division, Department of Agricultural Bio-Resources, National Academy of Agricultural Science, Rural Development Administration, Jeonju 55365, Korea
[6]Phyzen Genomics Institute, Seongnam 13558, Korea
[7]Crop Biotechnology Institute/GreenBio Science and Technology, Seoul National University, Pyeongchang 25354, Korea

ABSTRACT Miniature transposable elements (mTEs) such as miniature inverted-repeat transposable element (MITE), terminal repeat retrotransposon in miniature, and short interspersed element are exquisite sources for marker development. mTEs are short, non-autonomous and stably inherited. The high-copy members are widely distributed into the gene rich euchromatic regions. Here, we conducted a modified transposon display (TD) for a high-copy MITE family, BraSto-2 (Bs2). The Bs2-specific primers derived from conserved sequences of Bs2 members as well as *Mse*I adapter primers were used for polymerase chain reaction (PCR) in two *Brassica rapa* accessions, 'Chiifu' and 'Kenshin'. The pooled PCR products were sequenced by Illumina sequencing platform instead of high-resolution gel electrophoresis. Subsequent *in silico*-based insertion polymorphism (IP) analysis (next-generation sequencing [NGS]-based Bs2 transposon display) was conducted, which generated more than 99 putative polymorphic insertion sites between 'Chiifu' and 'Kenshin'. Among 90 successful PCR amplification, 34 showed Bs2 IP (IP-Bs2) between 'Chiifu' and 'Kenshin' accessions, 27 and seven 'Chiifu'- and 'Kenshin'-unique insertions, respectively. When the 90 IP-Bs2 primer sets were applied to 10 *Brassica* accessions, including four additional *B. rapa* and *B. oleracea* accessions, 69 (76%) showed insertion olymorphism among accessions. The IP-Bs2 were evenly distributed through all the chromosomes and provide rich polymorphism among various *B. rapa* and *B. oleracea* accessions demonstrating the usefulness of these markers for various genetic diversity and molecular breeding studies in *Brassica*. In addition, NGS-based TD will be applicable to various high copy transposable elements family for high throughput and rapid polymorphic marker development which will be helpful for efficient plant genomics and breeding purposes.

Keywords Transposon display, Next-generation sequencing, Insertion polymorphism, *Brassica*, BraSto-2

INTRODUCTION

Transposable elements (TEs) account for the largest fraction (up to 85%) of most plant genomes and play tremendous control on the genome function and evolution (Feschotte 2008; Arkhipova *et al*. 2012; Bire and Rouleux-Bonnin 2012). TEs are classified into either DNA transposon or retrotransposon based on their transposition mechanisms. Likewise, TEs can be grouped as either autonomous (aTEs) or non-autonomous (nTEs) depending on the presence or absence of functional genes for transposition, respectively (Wicker *et al*. 2007; Sampath and Yang 2014). The nTEs include large retrotransposon derivatives, terminal repeat retrotransposon in miniature

Received August 11, 2016; Revised August 20, 2016; Accepted August 23, 2016; Published August 31, 2016
***Corresponding author** Tae-Jin Yang, tjyang@snu.ac.kr, Tel: +82-2-880-4547, Fax: +82-2-8873-2056

(TRIM), short interspersed elements (SINEs), and miniature inverted-repeat transposable elements (MITEs) (Casa *et al*. 2000; Shedlock and Okada 2000; Witte *et al*. 2001). Due to their miniature structure (<1,000 bp) TRIM, SINEs, and MITEs are also referred to as miniature transposable elements (mTEs) (Wessler *et al*. 1995; Okada *et al*. 1997; Casacuberta and Santiago 2003; Feschotte and Pritham 2007). Important characteristics of mTE such as their ubiquity, stable inheritance, dispersed and high-copy presence in the genome, and close association with genic regions provide better opportunity for marker development (Sampath and Yang 2014; Sampath *et al*. 2015).

DNA markers are used in a wide range of genomic and breeding applications such as construction of genetic linkage maps, genomics assisted breeding, genome-wide association studies and evolutionary studies (Purugganan and Wessler 1995; Casa *et al*. 2000; Kwon *et al*. 2007; Yaakov *et al*. 2012; Varshney *et al*. 2013). DNA markers have been developed using various methods like random amplified polymorphic DNA, restriction fragment length polymorphism, simple sequence repeats, amplified fragment length polymorphism (AFLP), sequence characterized amplified region, and single nucleotide polymorphism (Agarwal *et al*. 2008; Kalendar *et al*. 2011; Varshney *et al*. 2013). Moreover, development of polymorphic markers between close relatives or same species are laborious and time consuming due to its high homologous nature. Combination of multiple marker type provides better genome coverage for genetic linkage map and association map (Agarwal *et al*. 2008).

TE-based molecular markers such as inter-retrotransposon amplified polymorphism, retrotransposon-microsatellite amplified polymorphism, sequence-specific amplification polymorphism, insertion polymorphism based on retrotransposon and DNA transposon, inter-MITE polymorphism and transposon display (TD) (Agarwal *et al*. 2008; Kalendar *et al*. 2011; Shirasawa *et al*. 2012) have been successfully applied for the various genomics purposes such as genetic diversity, inspection of clonal variation, identifying unambiguous gene flow between closely related species and breeding (Deragon and Zhang 2006; Bire and Rouleux-Bonnin 2012; Carrier *et al*. 2012). DNA polymorphisms are used to identify molecular markers for important agronomic traits controlled by single gene or quantitative trait loci (Monden *et al*. 2009; Kalendar *et al*. 2011; Fattash *et al*. 2013). TD is a modified AFLP method which target the transposon to detect TE insertion polymorphisms (Casa *et al*. 2000). Using traditional gel based TD analysis has lot of limitations to develop high quality marker due to high copy nature of the mTEs. Also gel based TD requires more time, professional skill to recover and sequence the polymorphic bands. Most importantly it requires multiple rounds of experiment to clearly amplify all or most of the mTE insertions (Casa *et al*. 2004; Kwon *et al*. 2007).

Next-generation sequencing (NGS) provides fast, accurate, and cost effective way to determine the order of nucleotide bases by parallel sequencing of DNA/RNA fragments which has wide range of application towards complete decoding and genomics research for crop improvement by advanced genotyping (Patel *et al*. 2015). It can be successfully applied for multiplexing with many accessions or population in a single step using barcode sequence as tags (Varshney *et al*. 2009; Wood *et al*. 2010; Davey *et al*. 2011; Zhang *et al*. 2011). Taking advantage of the ubiquitous and random distribution nature of the mTEs, we developed large scale markers for *B. rapa* genome using a high copy stowaway MITE family, BraSto-2 (Bs2) (Sampath *et al*. 2013). Bs2, a stowaway MITE family used for the display analysis was recently characterized and comparatively analyzed, and was found out to be present as high copy (500-1,500) in the *Brassica* genome (Murukarthick *et al*. 2014; Sampath *et al*. 2014). Here, we developed a TD for the Bs2 members by applying NGS sequencing to uncover the insertion polymorphism and develop large-scale polymorphic markers mediated by recent insertion polymorphism of the Bs2 members (IP-Bs2 markers) among *Brassica* accessions. The IP-Bs2 markers are clearly identified in agarose gel-based markers which can be applied for various molecular breeding purposes in *Brassica*.

Table 1. List of accessions used for the display and insertion survey.

No.	ID	Species	Accession no.	Reference
1	Br1	*Brassica rapa*	'Chiifu' (C)	(Wang *et al*. 2011)
2	Br2	*B. rapa*	'Kenshin' (K)	(Sampath *et al*. 2013)
3	Br3	*B. rapa*	OC 1	(Lee *et al*. 2014)
4	Br4	*B. rapa*	OC 2	(Lee *et al*. 2014)
5	Br5	*B. rapa*	YE 1	(Lee *et al*. 2014)
6	Br6	*B. rapa*	YE 2	(Lee *et al*. 2014)
7	Bo1	*Brassica oleracea*	C1234	(Lee *et al*. 2015)
8	Bo2	*B. oleracea*	C1184	(Lee *et al*. 2015)
9	Bo3	*B. oleracea*	C1235	(Lee *et al*. 2015)
10	Bo4	*B. oleracea*	C1176	(Lee *et al*. 2015)

MATERIALS AND METHODS

Plant materials and genomic DNA extraction

Genomic DNA from a total of ten accessions from *B. rapa* and *B. oleracea* were extracted using modified cetyltrimethylammonium bromide method (Allen *et al*. 2006) and the quality of the DNA were quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). High quality DNA was used for TD and insertion polymorphism survey (Table 1) (Lee *et al*. 2014, 2015).

NGS-based Transposon display of Bs2 MITE family

A high-copy MITE family, Bs2, was used for NGS-based TD analysis against two *B. rapa* accessions, 'Chiifu' and 'Kenshin', with some modifications from gel-based MITE display (Casa *et al*. 2004) (Fig. 1). Briefly, 500 ng of the genomic DNA was digested with a tetra-cutter *Mse*I at 37°C for 2 hours and the digested DNA was ligated with *Mse*I adaptor forward (5'-GACGATGAGTCCTGAG-3'), *Mse*I adaptor reverse (5'-TACTCAGGACTCAT-3') sequences using one unit of T4 ligase enzyme at 16°C for 3 hours. The ligated products were diluted to five-fold with sterile water then subjected to pre-selective amplification using primer specific to *Mse*I adaptor sequence (*Mse*I+0 5'-GACGATGAGTCCTGAGTA-3') and a Bs2 specific degenerative primer (Bs2 primer: 5'-CGACTTATAWT-AAAAAACGGAGGG-3') (Fig. 2). Degenerative primer was developed on the conserved sequence based on multiple sequence alignments of the Bs2 members. Pre-amplification reaction mixture (50 μl total) consisted of 10 μl ligated DNA, 1× polymerase chain reaction (PCR) buffer, 0.2 μM of each primer, 2.5 μM dNTPs, and 1 unit *Taq* DNA polymerase (Vivagen, Seongnam, Korea). PCR was carried out as 5 minutes at 94°C, 35 cycles of 95°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute, with a final 20-minute extension at 72°C, using ABI thermocycler (Applied Biosystems, Santa Clara, CA, USA). A 5 μl of pre-amplification PCR products were separated on 2% agarose gel, and the gels were stained with ethidium bromide and visualized on a UV transilluminator.

After the pre-amplification process, the products from 'Chiifu' and 'Kenshin' were purified using Qiagen PCR purification kit (Qiagen, Hilden, Germany). The purified PCR products were sent for sequencing by Illumina HiSeq2000 paired-end sequencing platform according to the manufactures protocol at Labgenomics, Seongnam, Korea (Fig. 1A). Briefly, the PCR products were pooled from 'Chiifu' and 'Kenshin' after labeling with two different Illumina barcode to perform the simultaneous multiplex sequencing using Illumina HiSeq2000 paired-end multiple sequencing method. The accession-specific reads were then extracted from the mixture based on the barcode information. Pair-end reads from 'Chiifu' and 'Kenshin' were mapped against the reference genome of *B. rapa* 'Chiifu' v1.2 to obtain the physical position. The redundant or duplicate sites were eliminated and the unique sites were then used for further analysis. The shared or common sites were manually identified based on physical position information and eliminated. The remaining candidate sites which are

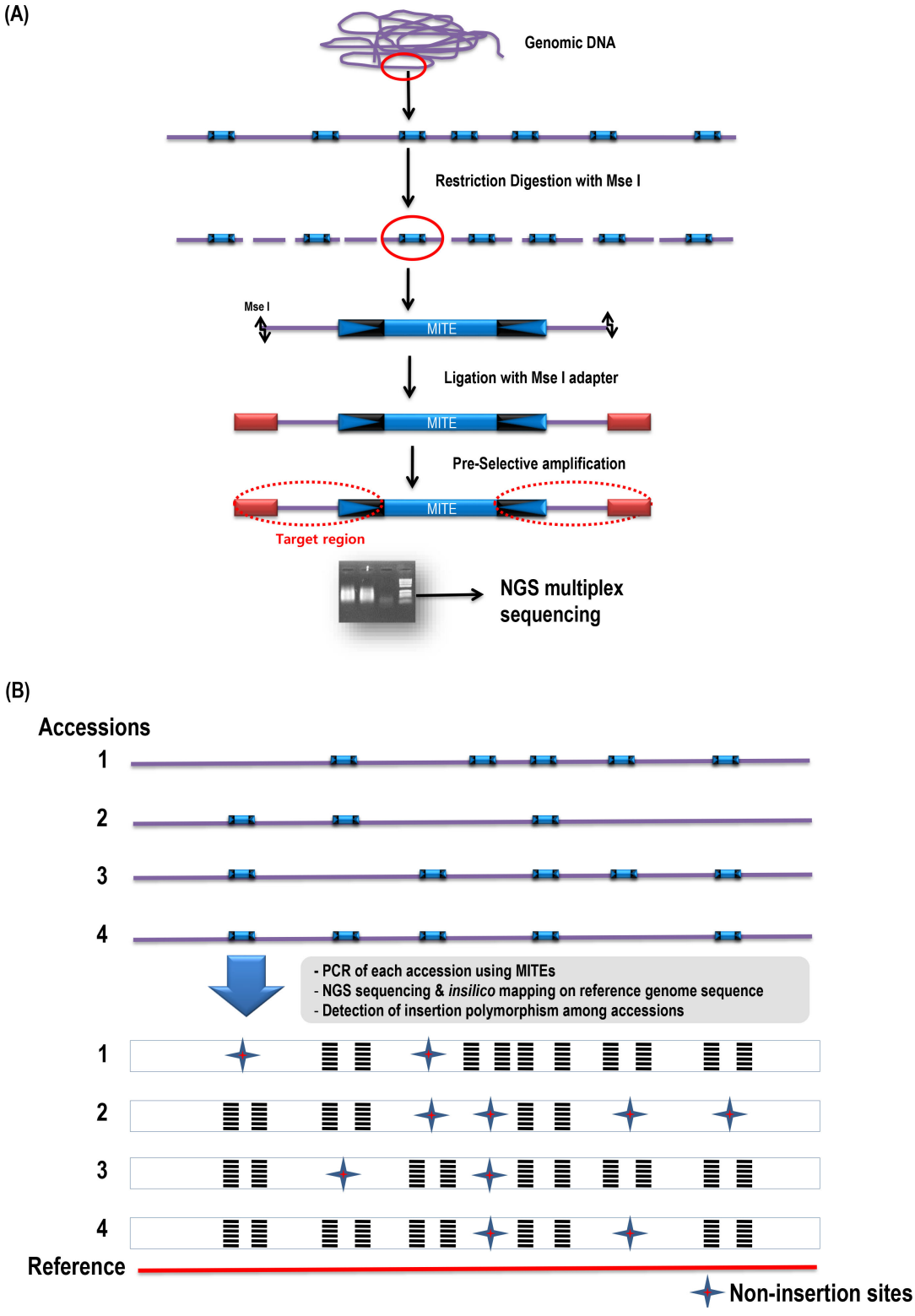

Fig. 1. Next-generation sequencing (NGS) based transposon display. (A) Steps involved in NGS-based transposon display analysis. The target region (red dotted circle) used for the sequencing. (B) Identification of polymorphism site (presence/absence of conserved miniature inverted-repeat transposable element [MITE] sequences) by analyzing reads from the different accessions. PCR: polymerase chain reaction.

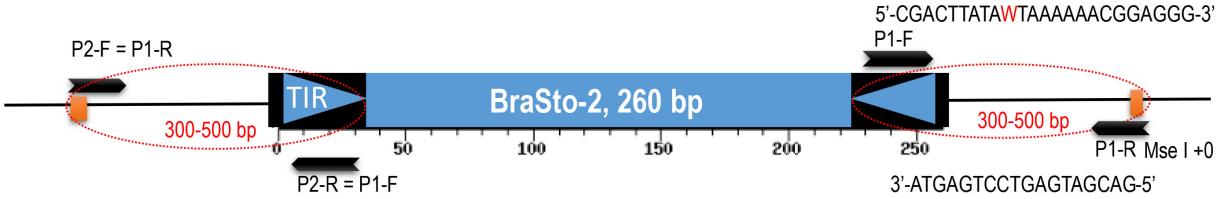

Fig. 2. Structure of the BraSto-2 (Bs2) miniature inverted-repeat transposable element (MITE) and primers for MITE-display. Primers from the Bs2 consensus region, terminal inverted repeat (TIR) and *Mse*I restriction enzyme site shown as arrows. Target regions are shown with dotted ovals. P: primers, F: forward, R: reverse. W in primer sequence is degenerate base symbol can bind to A/T.

Table 2. Summary of reads analysis from NGS-based transposon display of Bs2 MITE family against two *Brassica rapa* accessions.

Accession	Insertion sites based on *in silico* mapping		PCR validation		
	Total	Accession specific[z)]	Success	IP-Bs2 among Br1, Br2	IP-Bs2 among 10 accessions
Br1	127	83	75	27 (36)	59 (78)
Br2	60	16	15	7 (46)	10 (66)

Values are presented as number only or number (%).
[z)]Bs-2 sites specific to Br1 and Br2.
NGS: next-generation sequencing, MITE: miniature inverted-repeat transposable element, PCR: polymerase chain reaction, IP-Bs2: Insertion polymorphism of Brasto-2 (bs2) members.

predicted to be accession-specific were used for validation of polymorphic insertion analysis (Fig. 1B, Table 2).

Insertion polymorphisms analysis of Bs2 members (IP-Bs2)

In order to validate the accession-specific insertions, insertion polymorphisms were surveyed on 10 *Brassica* accessions including six *B. rapa* and four *B. oleracea* according to previous approach (Sampath *et al*. 2013). Briefly, PCR was carried out using the Bs2 flanking primers with the condition as 5 minutes at 94°C, 35 cycles of 95°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute, with a final 20-minute extension at 72°C, using ABI thermocycler. The 5 μl of PCR products were separated on 2% agarose gel, and the gels were stained with ethidium bromide and visualized on a UV transilluminator. The primers used for MIP analysis and polymorphisms information are listed in Table 3.

RESULTS

Development of NGS-based Bs2 transposon display

We have performed NGS-based TD for a high-copy MITE family, Bs2, using Illumina multiplex platform. It is a modification of a previous MITE display method reported by Casa *et al*. (2004), and developed using Bs2 against two *B. rapa* accessions, 'Chiifu' and 'Kenshin' (Fig. 1). A degenerate primer was developed from the most conserved region of the Bs2 by adding a degenerate nucleotide for the Bs2 specific primer (5'-CGACT-TATAWTAAAAAACGGAGGG-3'). The Bs2-specific primer binds to both the end and amplify the flanking regions for both side as well as amplify with the primer based on *Mse*I adaptor sequence (Fig. 2). Thus, we tried to find the insertion polymorphism of Bs2 members (IP-BS2) between *B. rapa* 'Chiifu' and 'Kenshin' accessions. NGS Pair-reads were derived from the Bs2 flaking regions were mapped on to the *B. rapa* 'Chiifu' pseudo-chromosome sequences (Wang *et al*. 2011). Reads showed 1-25×

Table 3. Insertion polymorphisms survey of Bs2 candidates based on NGS-based transposon display analysis against 10 *Brassica* accessions.

No.	Primer sequence		Product size (bp)	Temperature (°C)	Primer source			*In silico* specificity	PCR validation	Gel profile[z]									
	Forward (5'-3')	Reverse (5'-3')			Chr#	Start	End			Br1	Br2	Br3	Br 4	Br 5	Br 6	Bo1	Bo2	Bo3	Bo4
1	TTGTTTGATCAGGGGAGCAT	CCAGGGTACCTATCCGCTTT	858	58.45	1	21801279	21802374	C	C	1	2	-	2	2	2	2	2	2	2
2	TTGCAGATTTGATGTTGTGAA	GCGCGGGTTATTACCTAGTG	713	57	2	23504716	23506305	C	C	1	2	3	1	3	1	1	1	1	1
3	CTACCGCCAAGTTCAAGCA	TAAGGGACCCCAAAGCATAA	773	56.85	3	742954	744043	C	C	1	2	3	3	3	3	1	1	1	1
4	TTGAGCCTAGTCCGAGCAAT	CTCATCTCCAAACCCCATTC	749	58.4	3	10933621	10934710	C	C	1	2	1	3	3	1	1	1	1	1
5	ACCGTGGTGAGGTAAAAGGA	TGTTCTTTGCCTTGGAACCT	751	57.4	3	18816045	18817634	C	C	1	2	1	3	3	1	2	2	2	2
6	TGCAGATCTCTTTGCTCATCA	CTGTTCTCTGCGATGCTCAC	746	58.95	3	20935930	20936992	C	C	1	2	2	2	2	2	2	2	2	2
7	GCATCTCTGAGCTGGTTTCC	GTCCTCGTTGACGGAGAAAG	976	60.5	3	17138648	17140237	C	C	1	2	-	-	-	-	-	-	-	-
8	GCAAATTATGCACAATCTTACAA	TGGATATATGATGCTGTCAAAAA	751	55.7	3	23886171	23887233	C	C	1	2	-	2	1	2	-	-	-	-
9	GGAATCGAATGGGATCAAAA	TCTAAAAACGCTGGCTCCAT	886	55.35	5	11493271	11494360	C	C	1	2	3	3	3	3	-	-	-	-
10	CCGGCTGATTGCTCTAATGT	CAACATATGCCTCCACCACA	816	58.4	5	104661	105738	C	C	1	2	-	2	2	2	-	-	-	-
11	TGAAACAAAACGCTTTCTCG	TCAAAATAGTCACCAATCGGAGT	988	56.8	5	3281961	3283550	C	C	1	2	-	1	1	1	-	-	-	-
13	CAATGCAAGCCTCACGTATG	CTTATTGGCCATGCCTGACT	647	58.4	5	22026785	22027585	C	C	1	2	-	-	-	-	-	-	-	-
12	CGCAAAAGTGTACAAAATCTCAA	TGGAGTAGACCTGGCGGTAA	644	59	5	2018577	2019377	C	C	1	2	-	-	-	-	-	-	-	-
14	GGCACAACCAAGCCAATAAT	TTACACGCACCGAATTTGAC	735	56.4	6	22061648	22062736	C	C	1	2	2	2	2	2	2	2	2	2
15	AAACGGCAATTCGTCTTTTC	TTGCCTCGTAGCACTTTTCTC	755	56.85	6	18749148	18750231	C	C	1	2	-	3	2	2	2	2	2	2
16	AAAGAAAGCTTTGGCTTAGCTG	ACCCATATCACCCGACCATA	710	58.4	6	23123270	23124363	C	C	1	2	-	3	3	1	2	2	2	2
17	GAAGAAGCGAGCGAGAAGAA	CTTGCCTTCTGATCCCAATC	910	58.4	7	22424472	22426064	C	C	1	2	-	3	1	1	2	2	2	2
18	ATGTCGCAACTGAACCAAAA	CAAATTACATTCGGGGCCTA	723	55.35	8	3233906	3235467	C	C	1	2	2	3	3	2	-	-	-	-
19	CCAGCTCACCACTTCACAAA	CAAAACAATCGGTTGGGAAT	712	56.35	8	16701066	16702155	C	C	1	2	2	2	2	1	-	-	-	-
20	TGTACGTACGTGAGAATGAGATAAT	ACCCTCATGATGCATGGTTT	726	58.65	8	2388601	2389689	C	C	1	-	-	-	-	-	2	-	-	2
21	ATAATAGGCGGCAAGAGCAC	AAACCATCGAAATGCTCACC	705	57.4	9	3996150	3997235	C	C	3	2	2	3	3	2	-	-	-	-
22	CGATGGTACATCAAAAACAAACA	CCATATGGTCCAAGGAAGGA	824	57.95	9	23729398	23730487	C	C	1	2	3	3	1	2	2	2	2	2
23	AATTGGGACGAAAAGGGATT	CTTTCGGAAACAGAGGGTGT	768	56.35	9	22776	23864	C	C	1	2	-	2	2	2	-	-	-	-
24	AGCCTACCGCTTAATGCAAA	TGTACAATGTATTTTCCTAACCAAAG	778	58.2	9	6049964	6051065	C	C	1	2	-	1	1	1	-	-	-	-
25	ACAACGCACTTTCAAAAGCA	CACCGAAGTTTTCTTTTGCTG	814	55.85	9	6435580	6436680	C	C	1	2	-	-	-	-	-	-	-	-
26	TGAGAAGCGTTTTCTGAGCA	CGGGTGTTTTTATAAGTTACACGTT	827	58.65	10	15070338	15071427	C	C	1	2	1	3	1	1	2	2	2	2
27	CTCACCAGCAGGGACACATA	TGGGCCACATTTTCTTAGGT	805	58.45	10	15382535	15383623	C	C	1	2	-	3	3	2	2	2	2	2
28	TGTTTACGGCAAGAACAAGA	GGTGATCATGAAAGATGCAA	842	54.3	3	12382669	12383669	K	K	2	3	-	-	-	-	1	1	1	1
29	CACCTCCTTCTCGCAGTATT	GAGGAAGGAAAAGGTTCGAG	897	58.4	5	3096661	3097721	K	K	2	1	-	-	-	-	-	-	-	-
30	TTGGATCAGATTCGGTTTTT	ACGGCCAAAGATTACAACAT	795	53.3	5	4793442	4794518	K	K	2	3	1	1	1	1	-	-	-	-
31	CGAAACAAGAACCCAAAAAC	GCCCAATGACCACTCTAAAC	883	56.35	5	19979539	19980595	K	K	2	3	-	-	-	-	-	-	-	-
32	CCTTTGTGGGGTTTACTGTC	TCTGATTACCAAACCTTGCAT	843	56.95	6	17041904	17042957	K	K	2	1	-	-	-	-	2	2	2	2
33	GGTTTCCTTTGGTGGTGATA	AATAACCGGATCAAACCTGA	784	55.35	6	24368107	24369107	K	K	2	3	-	-	-	-	1	1	1	1
34	TGGTTGGAGATGGAAAATAAA	TCGAAATTCAGCTCAATCAA	808	52.9	9	27626655	27627536	K	K	2	1	2	2	2	2	2	2	2	2
35	GTGTAGCCTATGGGGACGAA	CGCCCCTAAAGACAGCTAAA	701	59	1	963071	964133	C	Shared	1	-	1	1	-	-	-	-	-	-
36	GCTGTCCCTGAAAAAGGAGA	TCCGGTTGGCTTAAAAATTG	843	56.35	1	13502760	13503861	C	Shared	3	3	3	3	3	3	1	1	1	1
37	AGCTCATACACCTCGGGAGA	ATGCAGCTCGTGTCTGAGAA	759	59.45	1	19445593	19446681	C	Shared	1	1	-	1	1	1	2	2	2	2
38	AGGTACGGTTTCTCGGATCA	CGTGCATAGCTGTAAAACGA	620	57.4	1	2584681	2585481	C	Shared	3	3	-	-	-	-	-	-	-	-
39	ACAGAGGACCATACCGGAAC	AAGTCCTAAATACTCCCTCCGTTT	694	61.15	1	2961011	2961811	C	Shared	1	1	-	-	-	-	-	-	-	-
40	TTCACCACGAGTTGTCTTCG	GCGTTTCTCAATTCCTCTGC	611	58.4	1	9276641	9277441	C	Shared	1	1	-	-	-	-	-	-	-	-
41	TTGCATTAGCCGACAATGAT	GGAGCTTCTTCTTTCCACCA	630	56.35	1	10697577	10698377	C	Shared	1	1	-	-	-	-	2	2	2	2
42	TGGGAATCGTAGGAATGGAG	CCATCAGAGTCTCACCGACA	619	59.45	1	14754268	14755068	C	Shared	1	1	-	-	-	-	1	1	1	1
43	TCAAGCAACTTCTCCAGCAA	AGGTCAGTGCAAGCAAAGGT	432	57.4	1	18275269	18276069	C	Shared	1	1	-	-	-	-	1	1	1	1
44	ATCCGATCAACATTGCCTTC	CATTTTGTGGTGAAGCTGGA	652	56.4	1	22118044	22118844	C	Shared	1	1	-	-	-	-	2	2	2	2
45	GAAGCCACTACTGGTGTGTATGA	TGACATCATTCTTGCTATGATCC	661	61.1	1	23625986	23626786	C	Shared	1	1	-	-	-	-	-	-	-	-
46	GCAATTGTCAACAATACGTAAGC	TGTCGTTTTCGGTTTTCTATG	659	57.4	1	27587965	27588765	C	Shared	3	1	-	-	-	-	-	-	-	-
47	TTGGGGACCAATACCAATGT	GAGTGTTGGCCTTCGTCTTC	857	58.45	2	8839647	8841236	C	Shared	1	1	1	1	1	1	2	2	2	2

Table 3. Continued.

No.	Primer sequence		Product size (bp)	Temperature (°C)	Primer source			*In silico* specificity	PCR validation	Gel profile[z]									
	Forward (5'-3')	Reverse (5'-3')			Chr#	Start	End			Br1	Br2	Br3	Br 4	Br 5	Br 6	Bo1	Bo2	Bo3	Bo4
48	AAAAGCCTAAGGGCATCTCC	AATGCCTGCCCGTTACTCTA	856	58.4	2	20825514	20826587	C	Shared	1	1	1	1	1	1	2	2	2	2
49	TGGTCTGATTGGTTCATTGG	CTGCAAAATAACCGGTTTGA	719	55.35	2	532895	533984	C	Shared	1	1	-	1	1	1	2	2	2	2
50	GCTGTTGATATCGAAGAATGTGA	AAAACCGGAAGGAGTAACAAAA	707	57.95	2	6905529	6906618	C	Shared	1	1	-	1	-	1	-	-	-	-
51	CCGTAGAATGTGTGGGTGAA	AGAAGGCAAAGGCAAAGACA	809	57.4	2	25921691	25922779	C	Shared	3	3	-	3	3	2	-	-	-	-
52	CAAAGCCAGCTTCGTCTTTC	TTTTGGAAACGAGGGAGTACA	831	57.9	2	27426840	27427899	C	Shared	1	1	-	1	1	1	-	-	-	-
53	CCAAGGGTGTTAGGGATATTT	CCCATACCTTTTCAAACCAG	812	56.9	2	7699471	7700522	K	Shared	1	1	-	-	-	-	-	-	-	-
54	CAGGTTGTTGTGGGTTTTGA	ACAGTCGCCATTTCTCACCT	705	57.4	3	2817737	2818831	C	Shared	1	1	1	1	1	1	-	-	-	-
55	AACGTGTGTGGGTGAAAGTG	TGTGTACATGGCATTTGCTG	886	57.4	3	24785163	24786251	C	Shared	1	1	2	2	2	2	-	-	-	-
56	CCCGATAAAAATTTATGGTAGCAC	ACGCAAGTCAGAGCTGGTTA	876	59.25	3	29733660	29734749	C	Shared	1	1	1	1	1	1	-	-	-	-
57	TGCTGCAAATGCAACTTTTT	CCTGCCCCAACTGTATTTTC	934	55.35	3	22824441	22825982	C	Shared	1	1	-	1	-	-	-	-	-	-
58	CGAATATGGACACGTGAAAA	GTCCATAGAGGCATCCAAAC	778	56.35	3	8150869	8151901	K	Shared	3	3	-	-	-	-	2	2	2	2
59	AAATGTCGCCACTGAATCTG	AACCGAATCAAACCAACCAG	869	56.4	4	15959699	15960723	C	Shared	1	1	1	1	1	1	2	2	2	2
60	TGAATTGAAGCCACAAGCTA	CACGTGTTGTTTTGTTCGTT	842	54.3	4	13346608	13347631	K	Shared	1	1	-	-	-	-	1	1	1	1
61	TGTCGTTTTGGTTTTCAATG	GTGCCAGATTTTTAGCGACT	802	54.3	4	15212718	15213750	K	Shared	3	3	-	-	-	-	1	1	1	1
62	AGCAAGTGCCTCTCGAGTCT	TCAAAATAGTCACCAATCGGAGT	851	59.9	5	3281461	3283050	C	Shared	1	1	1	1	1	2	-	-	-	-
63	TTACGGAGGGAAAGCAGAGA	CGTAAATGCTCTCCAAATGC	776	57.4	5	10244950	10246022	C	Shared	1	1	3	3	3	1	-	-	-	-
64	TCGATTTCTTCCCATCAACC	TTGGAAGTAGCTCCGCAAAT	883	56.4	5	12834818	12835901	C	Shared	3	3	3	3	3	3	2	2	2	2
65	CTCTTCCGCTCTACCAACTGA	TCCTTCAACCTCCAACATGA	891	58.85	5	18108358	18109422	C	Shared	1	1	1	1	1	1	1	1	1	1
66	CTGCTTGAATCGGCTACAAA	CGGGCATCCAAATACTCTGT	826	57.4	5	18163408	18164496	C	Shared	1	1	1	1	1	1	3	2	2	3
67	GGAATGGTGAAGGACCTGAA	CCTAGCTCGACCATGGAGAC	786	60.45	5	19371747	19372834	C	Shared	1	1	1	1	1	1	-	-	-	-
68	TTTGCACCTAATTGATTTCCTTT	TGTCACGTGTGAAACATACTCC	811	58	5	21905619	21906708	C	Shared	1	1	1	1	1	1	-	-	-	-
69	GGGTGGTTAACGAGCCAGTA	TGGAAAACCATGGCAAAAA	879	55.7	5	23188047	23189647	C	Shared	1	1	1	1	1	1	2	2	3	2
70	GCGTGGTTACCTTCAATTCC	AGACTCGAGAGGCACTTGCT	614	59.45	5	3282051	3282851	C	Shared	1	1	-	-	-	-	1	1	1	1
71	TTTTTGGAGATGCATTTAGTGG	ACGCCAAAACTGAAAAGGAA	465	55.45	5	7674505	7675305	C	Shared	1	1	-	-	-	-	1	1	1	1
72	GGAGTTGCCATATTGGAAGG	GCCTCATACAGGAGGTGAGC	683	60.45	5	12835307	12836107	C	Shared	1	1	-	-	-	-	1	1	1	1
74	CAAAATCCACCGTCAAACTG	CCATTCAACCCGCTGTTACT	454	57.4	5	16754371	16755171	C	Shared	2	3	-	-	-	-	3	3	3	3
75	CGTGCATAGCTGTAAAACGAC	TGCATCTGCTGCTTTCATTT	622	56.85	5	18109146	18109946	C	Shared	2	3	-	-	-	-	-	-	-	-
73	AGGCCAACACAATAGGATCG	AAACGGCTACCACATCCAAG	442	58.4	5	13116126	13116926	C	Shared	1	1	-	-	-	-	1	1	1	1
76	CACGGTTGTGTGACAGATTG	GCCTAGGCTAGTGACCTCCA	713	60.45	6	18132269	18133370	C	Shared	1	1	2	2	2	2	2	2	2	2
77	ACACTGTGCGGACAAAAATG	TCTTTCTGCAAACCCCTAGC	841	57.4	6	22195497	22196585	C	Shared	1	1	3	3	1	2	-	-	-	-
78	TAAGGTGGGCCGTAACGTAG	GTCTCCGATGAAACGATGCT	845	59.45	6	15006054	15007142	C	Shared	1	1	-	3	2	1	-	-	-	-
79	TTGGGATGACAAGGATTTCT	CGACAAGCACAGAGACAAAG	885	56.35	6	2249171	2250222	K	Shared	1	1	-	-	-	-	-	-	-	-
80	CACATGGAACCTTTCTCCTC	TATCGGGTAAAGCCAATGAT	774	56.35	6	20148620	20149708	K	Shared	1	1	-	-	-	-	1	1	1	1
81	GGGGTTAGAATCGTCCTTTT	TTCTTGCGTGTTGGTATCAC	858	56.4	6	21234008	21235039	K	Shared	1	3	-	-	-	-	2	2	2	2
82	TGTAGACTCCTCCCAACGTCT	GGAAGTGGTGGATGCTGTTT	810	59.85	8	13609744	13610827	C	Shared	1	1	1	3	3	1	-	2	-	-
83	GGCAGAGAGCAGTTTCGATT	CTTCTCTGCCCAAACCTACC	897	59.45	8	17589879	17591468	C	Shared	1	1	2	3	1	2	-	-	-	-
84	TTGAAAAGCAAACCCTTCTC	ATTTTGGTTGGTTCATACCG	779	54.3	8	19318537	19319537	K	Shared	1	1	-	-	-	-	-	-	-	-
85	GCACTGAGCTTTGTCATGGA	TCTCACGTCCCTCTCCATCT	742	59.45	9	6120652	6121741	C	Shared	1	-	-	-	-	-	3	2	3	3
86	CGAGGAATGTGGTGATGATG	CGAGGAATGTGGTGATGATG	810	58.4	9	6223231	6224320	C	Shared	1	1	-	-	-	-	-	-	-	-
87	AGCAGCTATTGCATGGTCAC	TTGAGCTCTATTGGCAAGCA	722	57.4	9	19954499	19955588	C	Shared	1	1	1	1	1	1	2	2	2	2
88	CCACCCTTCAAGCATCAAAT	AAACAGAGACAACGCTGCTG	876	57.4	9	26437989	26439077	C	Shared	1	1	3	3	1	2	3	1	3	2
89	GGTTTGCTACCCACAAAATCA	CGCTCCTCTATTCGGACACT	898	58.95	9	26805934	26807535	C	Shared	1	1	1	1	1	2	-	-	-	-
90	AGGCCCCAATTTCCCTATAA	CCTTTTTCACCATTTACATTCG	877	56.5	10	7137934	7139011	C	Shared	1	1	1	3	3	1	-	-	-	-

[z]Scores of gel profile were followed by previous report (Murukarthick *et al.* 2014) as 1, full site; 2, empty site; 3, full and empty site; and -, no amplification. Bold used as a representative for each group in Fig. 3.

Bs2: BraSto-2, Chr#: chromosome numbers, NGS: next-generation sequencing, PCR: polymerase chain reaction, C: 'Chiifu', K: 'Kenshin'.

coverage against the target MITE flanking regions. Accession-specific Bs2 insertion sites for *B. rapa* 'Chiifu' and 'Kenshin' were identified by extensive *in silico* and manual analysis. A total of 127 and 60 perfectly paired insertion sites have been obtained from *B. rapa* 'Chiifu' and 'Kenshin' accessions, respectively (Table 2). Comparative analysis of candidate sites based on the physical position information revealed 83 and 16 insertions that were predicted to be unique to 'Chiifu' and 'Kenshin' accessions, respectively and 44 insertions were common between both accessions. These accession-specific candidates were used for validation by designing specific primer combinations for each site to detect insertion polymorphism (Table 3).

Validation of the IP-BS2 markers

PCR validation analysis was done for all of the 99 putative Bs2 insertion polymorphic candidates identified from *in silico* analysis. Out of 83 putative 'Chiifu' specific insertions, 75 were successfully amplified in which 27 (36%) candidates have produced desirable polymorphic insertions between 'Chiifu' and 'Kenshin' while 59 (78%) was observed among 10 accessions (Fig. 3A, D). Similarly, out of 16 putative 'Kenshin'-specific targets, 15 were successfully amplified. Seven (46%) polymorphic insertions between 'Chiifu' and 'Kenshin' while 10 (66%)

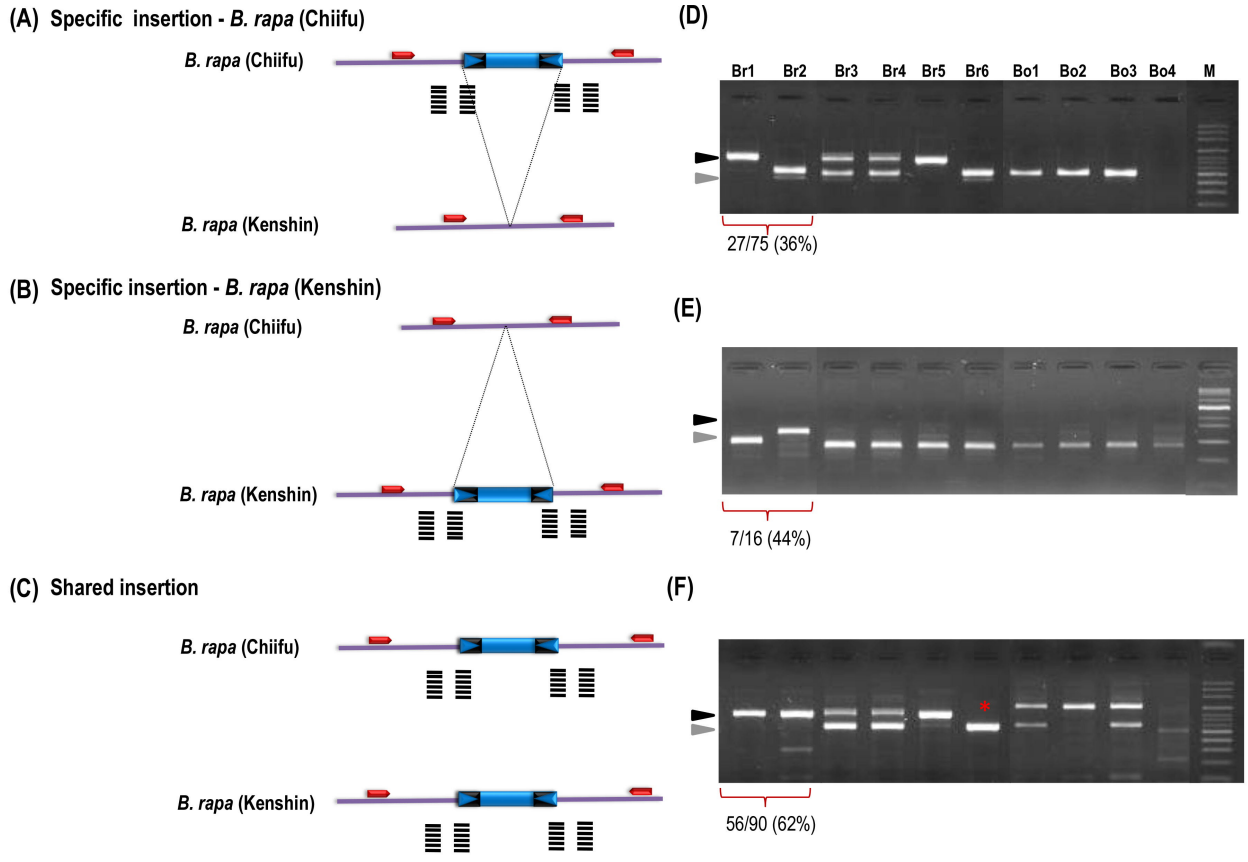

Fig. 3. Validation of sequencing based next-generation sequencing-based transposon display analysis of BraSto-2 (Bs2) miniature inverted-repeat transposable element (MITE) family members by insertion polymorphism survey. (A) Bs2 family member shows *Brassica rapa* 'Chiifu' specific insertion compare to 'Kenshin'. (B) Bs2 family member shows *B. rapa* 'Kenshin' specific insertion compare to 'Chiifu'. (C) Shared or common insertion Bs2 between 'Chiifu' and 'Kenshin'. Fig. 3D-F show the corresponding gel validation of A, B, and C, respectively. Black and grey arrow head indicate the MITE insertion (full site) and non-insertion (empty site), respectively. Star indicates the polymorphism in Br-6 produced by a shared insertion of 'Chiifu' and 'Kenshin'. Fig. 3D-F are based on the primers 3, 34, 88 from Table 3, respectively.

was observed among the 10 accessions (Fig. 3B, E). We observed 62% (56/90) of the insertions which were predicted to be polymorphic based on sequence analysis actually have a monomorphic pattern as shared insertions between the two accessions (Fig. 3C, F). In our approach we have identified 38% of the polymorphic insertions from a highly conserved MITE family in the *B. rapa* genome.

Genomic distribution of Bs2 and IP-BS2 markers in *B. rapa* genome

We surveyed genomic distribution of those polymorphic insertion on *B. rapa* 'Chiifu' reference genome. The *B. rapa* 'Chiifu' pseudo-chromosome sequences (version 1.2) contains 76 copies with 95:95 coverage of Bs2 members (hit with longer than 247 bp as 95% sequence homology) and 401 copies with 80:80 coverage (hit with longer than 208 bp as 80% sequence homology) Bs2 members. The 401 Bs2 members were shown on the *in silico* map which are distributed all over the chromosome regions. Among them, 207 (51%) were present within <1 kb vicinity of the gene (Sampath *et al*. 2013). *In silico* map with polymorphic insertion between 'Chiifu' and 'Kenshin' accessions show its random distribution (Fig. 4). Finding polymorphisms in random positions of the chromosome will be highly helpful for molecular breeding studies (Sampath *et al*. 2013).

DISCUSSION

NGS-based transposon display provides advantage for developing high-throughput insertion polymorphism markers

With the advent of NGS technology, simultaneous sequencing of more than one genome in a population is made possible using barcodes. Also, large scale sequencing, marker discovery, validation and assessment is possible for genomes with or without available high-quality reference genome information. Integration of NGS technology into the TD to develop NGS-based TD is not only time-saving but also produces stacks of information. Though MITEs

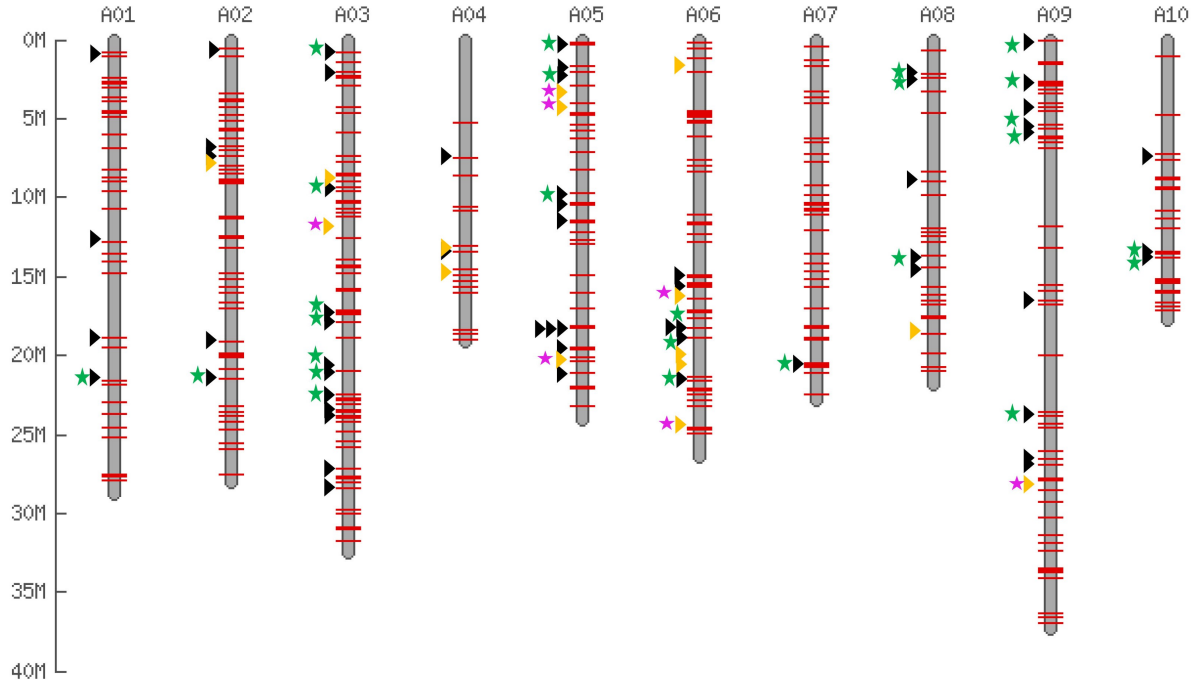

Fig. 4. *In silico* map of BraSto-2 (Bs2) members showing the surveyed and newly identified members on the *Brassica rapa* pseudo chromosome. Red bars represent the 401 Bs2 members on the *B. rapa* pseudo-chromosome. Black and yellow arrowheads indicate that the 90 *in silico* candidate members of Bs2 utilized by insertion polymorphism survey. Green and pink stars indicate the *B. rapa* 'Chiifu' and *B. rapa* 'Kenshin' specific insertions, respectively.

belong to Class II TEs (DNA transposons) it can amplify into hundreds of thousands of copies in a genome, which could be through positive selection or adaptive gap repair mechanisms or mobilization of autonomous partner element (Naito *et al*. 2009; Naito *et al*. 2014). MITE has been accumulated and amplified to high copies (> 22,000 in rice) in a genome (Naito *et al*. 2009; Naito *et al*. 2014). Likewise, Bs2 is present in a very high copy in the *B. rapa* genome (up to 500-1,500 copies) which is one of the highest copies among the 20 other MITE families in the present analysis (Sampath *et al*. 2015). Our recent analysis shows that Bs2 has differential amplification in copy numbers after *Brassica* speciation and up to now. But due to the high conservation of Bs2 in *B. rapa*, identification of polymorphic insertion is a very difficult task. Only 6% (3/50) polymorphic sites were identified between the *B. rapa* 'Chiifu' and 'Kenshin' accessions (Sampath *et al*. 2013).

Here, high-copy MITE family, Bs2, was analyzed by NGS-based TD against two *B. rapa* accessions 'Chiifu' and 'Kenshin' revealed that abundant polymorphic information (41%) suggest that the importance of NGS-based TD approach for high-throughput marker development. However, we could identify about 59% of the *in silico* candidates were shown shared insertion upon PCR validation. The discrepancy between sequence analysis which showed polymorphism between 'Chiifu' and 'Kenshin' and the actual PCR product may be due to lack of sequencing of the particular MITE member in the 'Kenshin' or 'Chiifu' genome. This error can be minimized by increasing the depth of the sequencing. Also, we found 39 insertions that were absent in the reference genome suggesting that the actual genome has more number of Bs2 insertions that were not included in the reference genome.

Our analysis showed that NGS-based TD will be a very useful method for high throughput MITE insertion polymorphic (MIP) marker development because the NGS analysis provides the flanking sequence information for MIP marker development in a short period of time (Table 3). Moreover, NGS-based TD approach will amplify most or all copies of multiple mTE families in a single analysis which will reduce the cost and time. More curation of data analysis like comparative analysis with more depth reads will increase the polymorphism ratio. This approach will be also effective to other TEs like TRIM and SINEs, and is also feasible for tandem repeats and any other conserved domains, which has a moderate to high copy number like centromeric tandem repeats, LRR-genes and R-genes. Furthermore, NGS-based TD has high advantage over conventional gel-based MITE display. Because, due to multiple number of bands, identification and development of markers through conventional gel based TD analysis is very difficult and time consuming and demands professional skill. Furthermore, TD requires multiple rounds of selective amplification in order to amplify all or most of the members of this high copy MITE family like Bs2 which cannot be amplified/visualized by a single gel analysis. However, NGS-based TD can overcome those limitations and can even be performed for multiple MITE families and for multiple number of accessions in a single analysis, emphasizing the importance of NGS-based TD for developing high-quality markers.

Application of IP-Bs2 makers for various molecular breeding purposes

The stable heritability, abundance and co-dominant nature of IP makers give them more advantage over other markers. Moreover, makers developed from IP approach have been used for various molecular breeding studies such as genetic diversity analysis, trait identification, and candidate gene analysis (Monden *et al*. 2009; Yaakov *et al*. 2012; Sampath *et al*. 2014). Our analysis based on the genome-specific MITE insertion showed high genetic diversity among various *B. rapa* accessions suggesting the importance of the MIP markers for the diversity analysis (Fig. 3D). Moreover, the markers which didn't show any polymorphisms between the 'Chiifu' and 'Kenshin' accession have produced polymorphisms between the other *B. rapa* accessions (Fig. 3F). This indicates that not only the 34 polymorphic markers but also the other 56 markers are highly valuable for various molecular applications such as high density genetic mapping, diversity and evolution studies as well as identification of the genetic components of germplasm with agronomically important traits to *B. rapa* and its relatives. This research validates the usefulness of NGS-based TD in high-throughput marker development in a

short period. Moreover, this study will provide new insights on effective utilization of mTEs for genomic studies.

ACKNOWLEDGEMENTS

This work was supported by Golden Seed Project (Centre for Horticultural Seed Development, No. 213003-04-4-SB430), Ministry of Agriculture, Food and Rural Affairs (MAFRA) and the Next Biogreen 21 program (PJ01101101) of the Rural Development Administration (RDA) in the Republic of Korea.

REFERENCES

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

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat. Protoc. 1: 2320-2325.

Arkhipova IR, Batzer MA, Brosius J, Feschotte C, Moran JV, Schmitz J, *et al*. 2012. Genomic impact of eukaryotic transposable elements. Mob. DNA 3: 19.

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

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

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

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

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

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12: 499-510.

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

Fattash I, Rooke R, Wong A, Hui C, Luu T, Bhardwaj P, *et al*. 2013. Miniature inverted-repeat transposable elements: discovery, distribution, and activity. Genome 56: 475-486.

Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. Annu. Rev. Genet. 41: 331-368.

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

Kalendar R, Flavell AJ, Ellis TH, Sjakste T, Moisy C, Schulman AH. 2011. Analysis of plant diversity with retrotransposon-based molecular markers. Heredity (Edinb) 106: 520-530.

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

Lee J, Izzah NK, Jayakodi M, Perumal S, Joh HJ, Lee HJ, *et al*. 2015. Genome-wide SNP identification and QTL mapping for black rot resistance in cabbage. BMC Plant Biol. 15: 32.

Lee S, Lee SC, Byun DH, Lee DY, Park JY, Lee JH, *et al*. 2014. Association of molecular markers derived from the *BrCRTISO1* gene with prolycopene-enriched orange-colored leaves in *Brassica rapa*. Theor. Appl. Genet. 127: 179-191.

Monden Y, Naito K, Okumoto Y, Saito H, Oki N, Tsukiyama T, *et al*. 2009. High potential of a transposon *mPing* as a marker system in *japonica x japonica* cross in rice. DNA Res. 16: 131-140.

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

Naito K, Monden Y, Yasuda K, Saito H, Okumoto Y. 2014. *mPing*: the bursting transposon. Breed. Sci. 64: 109-114.

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

Okada N, Hamada M, Ogiwara I, Ohshima K. 1997. SINEs and LINEs share common 3' sequences: a review. Gene 205: 229-243.

Patel DA, Zander M, Dalton-Morgan J, Batley J. 2015. Advances in plant genotyping: where the future will take us. Methods Mol. Biol. 1245: 1-11.

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

Sampath P, Lee J, Cheng F, Wang X, Yang TJ. 2015. Miniature transposable elements (mTEs): impacts and uses in the *Brassica* genome. p. 65-81. In: X. Wang, C. Kole (eds.). The *Brassica rapa* genome. Springer, Heidelberg.

Sampath P, Lee SC, Lee J, Izzah NK, Choi BS, Jin M, *et al*. 2013. Characterization of a new high copy *Stowaway* family MITE, *BRAMI-1* in *Brassica* genome. BMC Plant Biol. 13: 56.

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

Sampath P, Yang TJ. 2014. Miniature Inverted-repeat Transposable Elements (MITEs) as valuable genomic resources for the evolution and breeding of *Brassica* crops. Plant Breed. Biotech. 2: 322-333.

Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. Bioessays 22: 148-160.

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

Varshney RK, Mohan SM, Gaur PM, Gangarao NV, Pandey MK, Bohra A, *et al*. 2013. Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. Biotechnol. Adv. 31: 1120-1134.

Varshney RK, Nayak SN, May GD, Jackson SA. 2009. Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends Biotechnol. 27: 522-530.

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

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

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

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

Wood HM, Belvedere O, Conway C, Daly C, Chalkley R, Bickerdike M, *et al*. 2010. Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens. Nucleic Acids Res. 38: e151.

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

Zhang J, Chiodini R, Badr A, Zhang G. 2011. The impact of next-generation sequencing on genomics. J. Genet. Genomics 38: 95-109.