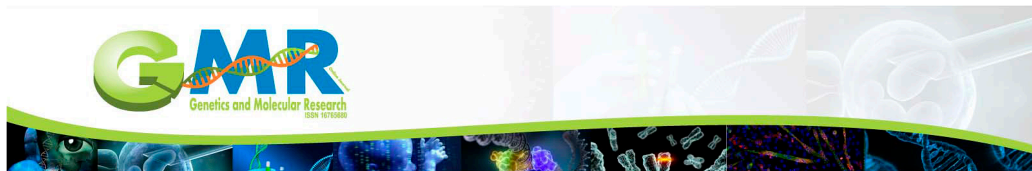

Development of a transposon-based marker system for mutation breeding in sorghum (*Sorghum bicolor* L.)

S.B. Im[1]*, S.-J. Kwon[1,2]*, J. Ryu[1], S.W. Jeong[1], J.B. Kim[1], J.-W. Ahn[1,2], S.H. Kim[1], Y.D. Jo[1], H.-I. Choi[1] and S.-Y. Kang[1]

[1]Advanced Radiation Technology Institute,
Korea Atomic Energy Research Institute, Jeongeup, Korea
[2]Radiation Biotechnology and Applied Radioisotope Science,
University of Science and Technology, Daejeon, Korea

*These authors contributed equally to this study.
Corresponding authors: S.-J. Kwon / S.-Y. Kang
E-mail: soonjaekwon@kaeri.re.kr / sykang@kaeri.re.kr

Genet. Mol. Res. 15 (3): gmr.15038713
Received April 18, 2016
Accepted June 6, 2016
Published September 16, 2016
DOI http://dx.doi.org/10.4238/gmr.15038713

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. Under certain circumstances, transposable elements (TE) can create or reverse mutations and alter the genome size of a cell. Sorghum (*Sorghum bicolor* L.) is promising for plant transposon tagging due to its small genome size and its low content of repetitive DNA. We developed a marker system based on targeted region amplification polymorphisms (TE-TRAP) that uses the terminal inverted repeats (TIRs) of transposons. A total of 3816 class 2 transposons belonging to the PIF/Harbinger family were identified from the whole sorghum genome that produced five primers, including eight types of TIRs. To define the applicability and utilization of TE-TRAP, we used 21 individuals that had been bred after γ-ray irradiation. In total, 31 TE-

TRAP, 16 TD, and 21 AFLP primer combinations generated 1133, 223, and 555 amplicons, respectively. The percent polymorphic marker was 62.8, 51.1, and 59.3% for the TE-TRAP, TD, and AFLP markers, respectively. Phylogenetic and principal component analyses revealed that TE-TRAP divided the 21 individuals into three groups. Analysis of molecular variance suggested that TE-TRAP had a higher level of genetic diversity than the other two marker systems. After verifying the efficiency of TE-TRAP, 189 sorghum individuals were used to investigate the associations between the markers and the γ-ray doses. Two significant associations were found among the polymorphic markers. This TE-based method provides a useful marker resource for mutation breeding research.

Key words: Sorghum; Transposable element; Mutation breeding; γ-ray; TE-TRAP

INTRODUCTION

Sorghum (*Sorghum bicolor* L.) is a C4 monocot species and one of the most important cereal crops. As a domesticated crop plant, it is used not only for human food, but also for fodder, animal feed, bio-energy, and as building material (Doggett, 1988; Rooney and Waniska, 2000; Lombardi et al., 2015). The complete genome sequence of sorghum was released for public use in early 2008, which enabled researchers to understand its complex traits at the DNA sequence level (Paterson et al., 2009). Sorghum has great potential for use in plant genomics due to its adaptation to harsh environments, the availability of a diverse germplasm collection, its small genome size, and the co-linearity of its genome with other cereal genomes (Bennetzen et al., 1998).

Mutation breeding is a useful method for crop improvement and it has played a critical role in sustainable agriculture. Mutation breeding using γ-rays directly produces mutant varieties without having to go through the otherwise lengthy and laborious process of conventional breeding. Irradiation has been successfully used for mutation breeding in various crops and ornamental plants and has proven to be a promising means of producing new genetic variants (Hara et al., 2003; Song and Kang, 2003).

Transposable elements (TEs) were first discovered by Barbara McClintock nearly 60 years ago using classical genetics in maize (McClintock, 1950). TEs are ubiquitous components of almost every investigated eukaryotic genome and repetitive sequences can change their location within a genome. TEs contribute significantly to the size, structure, and plasticity of genomes and also play an active role in genome evolution by helping their hosts adapt to new conditions by conferring useful characteristics (Zhang and Wessler, 2004; Wessler, 2006). TEs are categorized into two classes: class 1 elements, or retrotransposons, which move using a copy-and-paste mechanism to amplify intermediate RNA and class 2 elements, or DNA transposons, which move within the genome by excising from their original location to a new region using a cut-and-paste mechanism (Feschotte et al., 2002). Class 2 TEs are categorized into several subfamilies (Kunze et al., 1997). Among these, miniature inverted-repeat transposable elements (MITEs) are characterized by their small size (usually less than 500 bp), their lack of coding capacity, and their tendency to have short terminal

inverted repeats (TIRs) (Feschotte et al., 2002). MITEs are often discovered close to or within genes where they may affect gene expression (Santiago et al., 2002). MITEs can be further categorized into superfamilies, which include five discovered and characterized superfamilies among plant genomes (Feschotte and Pritham, 2007).

Among polymerase chain reaction (PCR)-based DNA marker systems, amplified fragment length polymorphism (AFLP), transposon display (TD), and targeted region amplification polymorphism (TRAP) are commonly and extensively used tools for assessment of variability in crops and genetic resources. AFLP analysis is a DNA fingerprinting technique developed by Vos et al. (1995). AFLP markers are dominant markers that detect high amounts of polymorphism (Mueller and Wolfenbarger, 1999). TD is a modified version of the conventional AFLP technique that has been applied successfully to a conserved sequence motif of a MITE element in maize (Casa et al., 2000). In this study, we developed a TE-TRAP marker system as a modification of the conventional TRAP marker system. The TE-TRAP marker system was developed to select a group of initial mutant plants after γ-ray irradiation. TE-TRAP is a simple PCR-based marker technique, which uses fixed primers from class 2 MITE TE superfamily sequence information and arbitrary primers that target intron or exon regions with an AT- or GC-rich core to amplify DNA fragments.

Previously, miniature Ping sequences in rice, which contain a putative P instability factor (PIF) family, were activated using γ-rays (Kikuchi et al., 2003). TE activity can be induced by environmental factors and, in particular, by physical stressors, such as γ-rays. For this reason, we conducted γ-ray irradiation of sorghum.

The objective of our study was to evaluate three different marker systems (TE-TRAP, TD, and AFLP) for genetic diversity, polymorphisms, and genetic distance, and to confirm the applicability and value of the TE-TRAP marker system developed in this study. We conducted an experiment to investigate the association between markers and γ-ray doses and to show that the TE-TRAP marker system is suitable for mutation selection.

MATERIAL AND METHODS

Plant materials and DNA extraction

Nine sorghum cultivars from Korea, the United States, and South Africa (Table 1) were used in this study. The 1000 dry seeds of each cultivar were irradiated by four different dose (100, 200, 300, 400 Gy) γ-rays emitted from a [^{60}Co] source at the Korea Atomic Energy Research Institute (KAERI). The seeds were planted in plots (3 x 6 m) and row spacing of 20 and 60 cm, respectively. After a month of germination, fresh leaf materials were harvested from five individuals each treated dose and original cultivars.

The genomic DNA was extracted from leaves of M_1 plant and each control using a modified cetyl trimethylammonium bromide method protocol (Doyle and Doyle, 1987). Genomic DNA was quantified using a NanoDrop ND-1000 spectrophotometer instrument (Thermo Fisher Scientific, Waltham, MA, USA) and it was normalized to a uniform concentration (100 ng/μL).

Sequence analysis and primer design

Whole sorghum genome sequences (Plant MITE databases; http://pmite.hzau.edu.cn)

was employed to identify class 2 MITEs using MITE Digger program (Yang, 2013). The confirmed MITEs were compared with the P-MITE database (http://pmite.hzau.edu.cn) for classification into TE super families. The compositions of five categories of MITEs (CACTA, hAT, PIF/Harbinger, Tc1/Mariner, Mutator) are shown in Figure 1. Using the PIF/Harbinger element sequence, eight types of TIRs were discovered and were aligned and matched in pairs with degenerate sequences. Sb_PIF primers were designed to match a 19-20 bp size corresponding to the 20-bp TIR of the PIF/Harbinger element. A *Zea mays* (Zm) primer, Zm_PIF_1, was designed using the TIR sequences of *Z. mays* mPIF deposited in the GenBank database (accession No. AF416298-AF416329). Isaac maps were designed from consensus sequences obtained from the GenBank database using basic information obtained from Lee et al. (2005).

Table 1. Cultivar name and origin of *Sorghum* germplasm used in this study.

No.	Cultivar name	Origin
1	DINE-A-MITE	Unknown
2	IS645	United States
3	IS2868	South Africa
4	Moktak	Korea (breeding cultivar)
5	Banwoldang	Korea (breeding cultivar)
6	Chal II	Korea (native species)
7	KLSO79125	Korea (native species)
8	KLSO79075	Korea (native species)
9	Mesusu	Korea (native species)

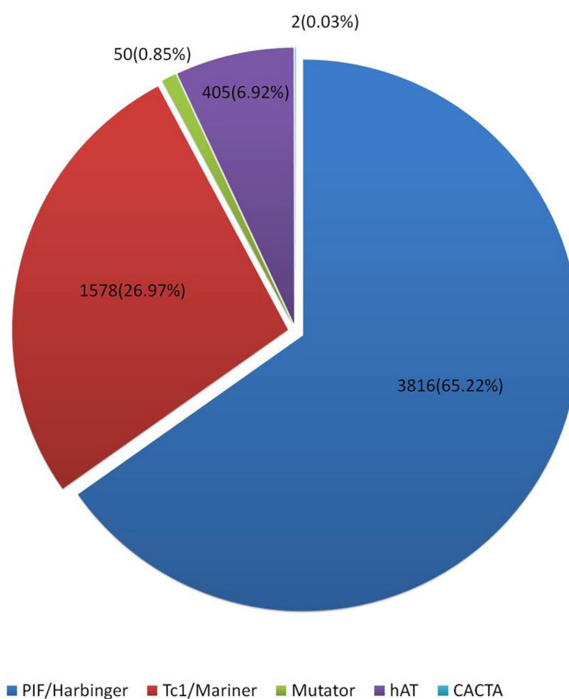

Figure 1. Distribution of class 2 transposable elements in the sorghum (*Sorghum bicolor*) whole genome sequence.

Generation of markers

The TE-TRAP, TD, and AFLP marker systems were applied to 21 individuals (five individuals each treated dose and a control) of the sorghum variety IS2868, which were irradiated with five ɤ-ray doses (0, 100, 200, 300, and 400 Gy). TE-TRAP amplification was carried out using a modified version of the protocol from Hu et al. (2005), which included seven fixed and six arbitrary primers. Arbitrary primers were used to target intron or exon regions with an AT- or GC-rich core to amplify the DNA fragments (Li and Quiros, 2001). The arbitrary primers used in this study were based on those described by Hu et al. (2005). A total volume of 20 μL was used for PCR amplification, which contained 2 μL genomic DNA (10 ng/μL), 1 μL fixed (10 pmol/μL) and arbitrary primer (2 pmol/μL), 0.4 μL dNTPs (10 mM), 2 μL 10X buffer, and 0.3 μL *Taq* polymerase (5 U/μL; Takara Ex *Taq*, Takara, Japan). The PCR amplification was performed by initially denaturing template DNA at 94°C for 2 min, followed by 5 cycles at 94°C for 45 s, 35°C for 45 s, and 72°C for 90 s, then 35 cycles at 94°C for 45 s, 50°C for 45 s, and 72°C for 90 s, and a final extension at 72°C for 7 min to terminate the reaction.

The TD and AFLP marker systems are similar. AFLP was conducted using two endonucleases (*Mse*I and *Eco*RI), whereas only one endonuclease (*Mse*I) was used for TD. The DNA was fully digested with the endonuclease at 37°C overnight and the reaction was terminated at 70°C for 2 min. The DNA fragments were ligated to adapters by adding 1 μL T4 DNA ligase to a total volume of 50 μL at 20°C for 2 h. Ligated DNA was diluted 10-fold and used for pre-amplification. Pre-amplification was done with the KRMP-0 and the Isaac map primer. PCR performed with one cycle at 72°C for 2 min and at 94°C for 3 min; 25 cycles were run at 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min; a final extension was run at 72°C for 5 min to complete the reaction. For selective amplification, the pre-amplified products were diluted 50-fold. The selective amplification consisted of 1 cycle at 94°C for 5 min; 10 "touchdown" cycles of 94°C for 30 s, 64°C for 30s, and 72°C for 1 min with a decrease in annealing temperature by 1°C each cycle; 28 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min; and 72°C for 5 min to terminate the reaction. Primer sequence information of each marker system is shown in Table 2. The amplified products were separated using a caliper LabChip GXII instrument (Caliper Life Sciences, USA) and the collected images were scored manually.

Data analysis

Each amplified fragment was scored with a binary code (1 or 0 for presence or absence, respectively). Based on the 0/1-matrix, we calculated gene diversity (GD), the percentage of monomorphic and polymorphic markers, polymorphic information content (PIC) (Nei, 1972; Yu et al., 2003), and genetic distance, using the genetic analysis package Power Marker (Liu and Muse, 2005). In addition, genetic differentiation was tested using the F_{ST} statistic, which estimates pairwise comparisons.

To compare the genetic differentiation of the three marker systems, we used a phylogenetic tree and principal component analysis (PCA). The phylogenetic tree was constructed using the unweighted pair group method with arithmetic mean (Sneath and Sokal, 1973), using the MEGA 6.06 program (Tamura et al., 2007), which is embedded in Power Marker. The PCA was used to classify and discriminate ɤ-ray irradiated sorghum individuals,

using the XLSTAT software (Addinsoft, 2008, USA). We then conducted analysis of molecular variance (AMOVA) (Schneider et al., 2000) using GenALEx v. 6.501 (Peakall and Smouse, 2006), to support the genetic differentiation information (obtained in the phylogenetic tree and PCA). The AMOVA estimates and separates total molecular variance among and within populations and then examines the significance of separated variance components using permutational testing procedures.

Table 2. Primer names and sequences for each of the three marker systems (TE-TRAP, AFLP, TD) used in the current study.

TE-TRAP			
Fixed primer	Sequence (5'-3')[1]	Arbitrary primer	Sequence (5'-3')
Sb_PIF_1a	KYY TGC ATT GTG AGT GCC CT	odd15	GCG AGG ATG CTA CTG GTT
Sb_PIF_1b	YYT GCA TTG TGA GTG CCC T	odd26	CTA TCT CTC GGG ACC AAA C
Sb_PIF_2	MYA TGC ATT GAG ACT GGC CT	sa4	TTA CCT TGG TCA TAC AAC ATT
Sb_PIF_3	ACW ACA TCC AAA CAA GGC CT	sa12	TTC TAG GTA ATC CAA CAA CA
Sb_PIF_4	TTS SGA ACT AAA CAA GGC CT	ga3	TCA TCT CAA ACC ATC TAC AC
Zm_PIF_1	RVK AAA CAA AYG GGV YC	ga5	GGA ACC AAA CAC ATG AAG A
Isaac Map	ATA GGG TGC GAT TCC GGT AGT G		
AFLP			
Adaptor	Sequence (5'-3')	Selective primer	Sequence (5'-3')
*Eco*RI/F	CTC GTA GAC TGC GTA CC	E/AA	GAC TGC GTA CCA ATT CAA
*Eco*RI/R	AAT TGG TAC GCA GTC	E/TG	GAC TGC GTA CCA ATT CTG
*Mse*I/F	GAC GAT GAG TCC TGA G	E/TT	GAC TGC GTA CCA ATT CTT
*Mse*I/R	TAC TCA GGA CTC AT	M/CAA	GAT GAG TCC TGA GTA ACA A
Pre-selective primer	Sequence (5'-3')	M/CAC	GAT GAG TCC TGA GTA ACA C
*Eco*RI	GAC TGC GTA CCA ATT C	M/CAG	GAT GAG TCC TGA GTA ACA G
*Mse*I	GAT GAG TCC TGA GTA A	M/CAT	GAT GAG TCC TGA GTA ACA T
		M/CTA	GAT GAG TCC TGA GTA ACT A
		M/CTC	GAT GAG TCC TGA GTA ACT C
		M/CTG	GAT GAG TCC TGA GTA ACT G
		M/CTT	GAT GAG TCC TGA GTA ACT T
TD			
Adaptor	Sequence (5'-3')	*Mse*I anchor	Sequence (5'-3')
KRMA-1	GAC GAT GAG TCC TGA G	KRMP-0	GAC GAT GAG TCC TGA GTA A
KRMA-2	TAC TCA GGA CTC AT	KRMP-GA	GAC GAT GAG TCC TGA GTA AGA
Selective primer	Sequence (5'-3')	KRMP-GAA	GAC GAT GAG TCC TGA GTA AGA A
Sb_PIF_1a	KYY TGC ATT GTG AGT GCC CT	KRMP-GAC	GAC GAT GAG TCC TGA GTA AGA C
Sb_PIF_1b	YYT GCA TTG TGA GTG CCC T	KRMP-GAG	GAC GAT GAG TCC TGA GTA AGA G
Sb_PIF_2	MYA TGC ATT GAG ACT GGC CT	KRMP-GAT	GAC GAT GAG TCC TGA GTA AGA T
Sb_PIF_3	ACW ACA TCC AAA CAA GGC CT	KRMP-CAC	GAC GAT GAG TCC TGA GTA ACA C
Sb_PIF_4	TTS SGA ACT AAA CAA GGC CT	KRMP-CAG	GAC GAT GAG TCC TGA GTA ACA G
Zm_PIF_1	RVK AAA CAA AYG GGV YC	KRMP-CAT	GAC GAT GAG TCC TGA GTA ACA T
Isaac map	ATA GGG TGC GAT TCC GGT AGT G	KRMP-CGT	GAC GAT GAG TCC TGA GTA ACC T
		KRMP-CTA	GAC GAT GAG TCC TGA GTA ACT A

The TE-TRAP marker system was used to analyze the association mapping between the markers and 189 individuals (nine accessions with five individuals of each ɤ-ray dose and each original cultivars), which were bred after ɤ-ray-irradiation, using the TASSEL 5.2.9 software (Bradbury et al., 2007). The associations between the markers and the ɤ-ray doses were tested using general linear model (GLM) and mixed linear model (MLM) with K-matrix (MLM + K) methods.

RESULTS

Identification of class 2 TEs and primer design

All five MITE superfamilies belonging to the class 2 TEs were identified in sorghum whole genome sequences (Figure 1). PIF/Harbinger elements were the most numerous

elements (65.20%) and TC1/Mariner elements were the second most numerous (27.00%). The number of hAT, Mutator, and CACTA elements was 6.90%, 50 (0.87%), and 2 (0.03%), respectively. To enrich the amplicons containing the flanking sequences of the PIF/Harbinger transposon, we designed Sb_PIF primer sequences using the outermost 20-bp TIR sequences of the PIF/Harbinger elements.

Analysis of genetic diversity

TE-TRAP was carried out with 31 primer combinations and a total of 1133 fragments were amplified. These fragments contained 20-50 amplicons ranging in size from 100 to 1500 bp per primer combination. Among these fragments, 421 (37.2%) were monomorphic and 712 (62.8%) were polymorphic (Table 3). For each primer combination, an average of 37 fragments was scored, and 23 were polymorphic. The highest level of polymorphism (85.4%) was obtained from primer combination Sb_PIF_3 + Sa12, whereas the lowest (43.2%) was obtained from primer combination Zm_PIF_1 + Sa12 (Table 4). Genetic diversity and PIC values ranged from 0.124 (Isaac map + sa12) to 0.335 (Sb_PIF_1a + sa4) and from 0.102 (Isaac map + sa12) to 0.264 (Sb_PIF_1a + sa4), with averages of 0.214 and 0.172, respectively.

Table 3. Gene diversity and polymorphic information content (PIC) of each marker system (TE-TRAP, TD, and AFLP).

	TE-TRAP	TD	AFLP
Number of markers	31	15	21
Number of polymorphic markers	712	114	329
Percent polymorphic markers	62.8	51.1	59.3
Number of monomorphic markers	421	109	226
Percent monomorphic markers	37.2	48.9	40.7
Gene diversity	0.214	0.148	0.159
PIC	0.172	0.122	0.132

We carried out the TD amplification with 16 primer combinations that resulted in a total of 223 amplicons. The number of amplicons amplified by each primer combination ranged from six (Sb_PIF_1b + TGC) to 26 (Sb_PIF_3 + GAT) and the sizes of the amplified fragments ranged from 100 to 1500 bp. A total of 223 fragments were scored, 109 (48.9%) of these were monomorphic, whereas 114 (51.1%) were polymorphic (Table 3). For each primer combination, on an average, 14 fragments were scored and seven were polymorphic. The highest level of polymorphism (68.8%) was obtained with primer combination Sb_PIF_2 + GAC, and the lowest level (33.3%) was obtained with primer combination Sb_PIF_1b + TGC and Isaac map + GAA (Table 4). The GD and PIC values ranged from 0.054 (Isaac map + GA) to 0.224 (Sb_PIF_3 + GA) and from 0.050 (Isaac map + GA) to 0.179 (Sb_PIF_3 + GA), averaging 0.148 and 0.122, respectively.

Amplification of ɤ-ray-irradiated sorghum sequences using the AFLP marker system generated multiple amplicons and abundant polymorphic fragments. The amplified fragments ranged from 100 to 1500 bp and the number of scorable fragments amplified by the 21 primer combinations ranged from 18 (TG + CTC) to 43 (TT + CAT). A total of 555 amplicons were scored of which 329 fragments (59.3%) were polymorphic and 226 fragments (40.7%) were monomorphic (Table 3). For each primer combination, an average of 26 fragments was scored and 16 were polymorphic. The highest level of polymorphism (86.8%) was obtained with

Table 4. Total fragments, number of polymorphic fragments and percentage of polymorphic fragment in each primer combination of marker systems.

Primer combination	Total fragments	Polymorphic fragments (%)	Primer combination	Total fragments	Polymorphic fragments (%)
TE-TRAP					
Sb_PIF_1a + odd15	42	23 (54.8)	Sb_PIF_3 + ga5	48	31 (64.6)
Sb_PIF_1a + odd26	38	22 (57.9)	Sb_PIF_3 + sa4	40	33 (82.5)
Sb_PIF_1a + ga5	37	27 (73.0)	Sb_PIF_3 + sa12	41	35 (85.4)
Sb_PIF_1a + sa4	40	34 (85.0)	Sb_PIF_4 + odd15	29	19 (65.5)
Sb_PIF_1a + sa12	27	17 (63.0)	Sb_PIF_4 + odd26	29	21 (72.4)
Sb_PIF_1b + odd15	40	23 (57.5)	Sb_PIF_4 + ga3	43	33 (76.7)
Sb_PIF_1b + odd26	38	21 (55.3)	Sb_PIF_4 + ga5	42	22 (52.4)
Sb_PIF_1b + ga5	42	23 (54.8)	Sb_PIF_4 + sa4	45	25 (55.6)
Sb_PIF_1b + sa12	37	22 (59.5)	Zm_PIF_1 + odd15	31	18 (58.1)
Sb_PIF_2 + odd15	40	25 (62.5)	Zm_PIF_1 + odd26	24	14 (58.3)
Sb_PIF_2 + odd26	28	14 (50.0)	Zm_PIF_1 + ga5	25	19 (76.0)
Sb_PIF_2 + ga3	28	21 (75.0)	Zm_PIF_1 + sa12	37	16 (43.2)
Sb_PIF_2 + ga5	39	21 (53.8)	Isaac map + odd26	28	17 (60.7)
Sb_PIF_2 + sa12	32	21 (65.6)	Isaac map + ga5	28	19 (67.9)
Sb_PIF_3 + odd15	48	32 (66.7)	Isaac map + sa12	40	18 (45.0)
Sb_PIF_3 + ga3	47	26 (55.3)			
AFLP					
AA + CAA	31	23 (74.2)	TG + CTA	20	13 (65.0)
AA + CAC	38	33 (86.8)	TG + CTC	18	9 (50.0)
AA + CAG	31	16 (51.6)	TG + CTT	19	13 (68.4)
AA + CAT	28	16 (57.1)	TT + CAA	19	7 (36.8)
AA + CTA	22	14 (63.6)	TT + CAC	32	17 (53.1)
AA + CTC	23	15 (65.2)	TT + CAG	22	10 (45.5)
AA + CTG	29	23 (79.3)	TT + CAT	43	31 (72.1)
TG + CAA	22	13 (59.1)	TT + CTA	24	16 (66.7)
TG + CAC	27	15 (55.6)	TT + CTC	28	11 (39.3)
TG + CAG	29	10 (34.5)	TT + CTG	30	11 (36.7)
TG + CAT	20	13 (65.0)			
TD					
Sb_PIF_1b + CAC	10	6 (60.0)	Sb_PIF_2 + GTA	10	6 (60.0)
Sb_PIF_1b + CAG	11	8 (72.7)	Sb_PIF_3 + GA	22	13 (59.1)
Sb_PIF_1b + CTA	9	5 (55.6)	Sb_PIF_3 + GAA	11	5 (45.5)
Sb_PIF_1b + GAG	7	3 (42.9)	Sb_PIF_3 + GAT	26	12 (46.2)
Sb_PIF_1b + TGC	6	2 (33.3)	Isaac map + GA	11	4 (36.4)
Sb_PIF_2 + GAA	19	11 (57.9)	Isaac map + GAA	15	5 (33.3)
Sb_PIF_2 + GAC	16	11 (68.8)	Isaac map + GAG	22	10 (45.5)
Sb_PIF_2 + GCT	12	7 (58.3)	Isaac map + GCT	16	6 (37.5)

primer combination AA + CAC, and the lowest level (34.5%) was obtained with primer combination TG + CAG (Table 4). The GD and PIC values ranged from 0.085 (TT + CTC) to 0.256 (TG + CTT) and from 0.073 (TT + CTC) to 0.207 (AA + CTG), respectively.

Genetic differentiation

Based on Nei's genetic distance (Nei, 1972), a phylogenetic tree was constructed based on genetic difference among 21 individuals. Based on genetic difference, the TE-TRAP, TD, and AFLP marker systems produced three, three, and two major groups, respectively. The PCA provided an alternative view of the genetic difference among the 21 individuals. The three groups produced by the TE-TRAP tree clustered within a genetic distance of 0.11. Group 1 comprised two individuals of 200 Gy and one individual of 300 Gy, whereas group 2 included two individuals of 200 Gy, three individuals of 300 Gy, and three individuals of 400 Gy. The third group contained one control individual, four individuals of 100 Gy, two individuals of 300 Gy, and two individuals of 400 Gy. The phylogenetic analysis corresponded very well with the PCA for the TE-TRAP marker system (Figures 2 and 3). In contrast, the phylogenetic tree and PCA results based on the TD and AFLP marker systems did not agree, due to unstable groupings produced by the PCA (data not shown).

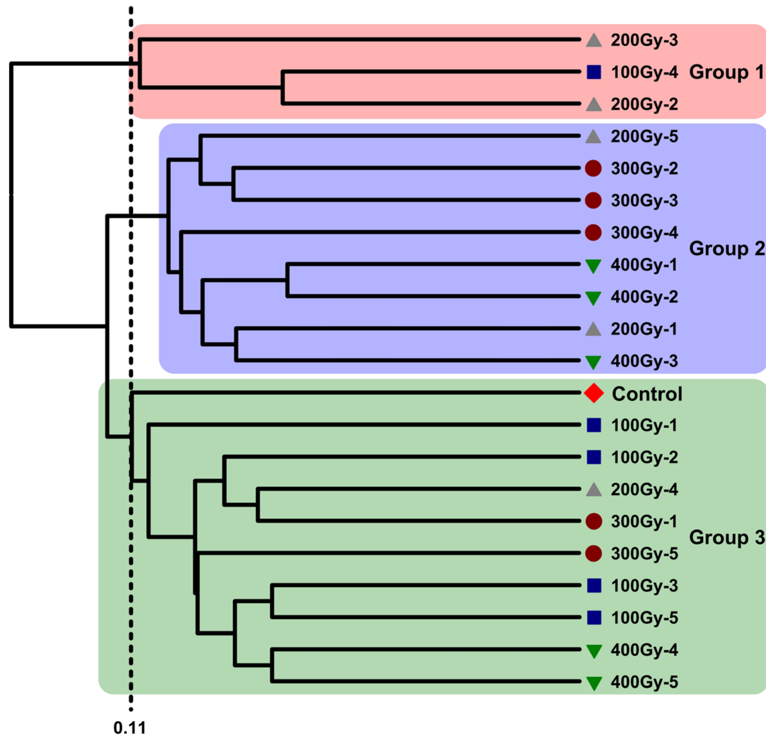

Figure 2. Phylogenetic tree of 21 ɣ-ray irradiated sorghum (*Sorghum bicolor*) plants using a TE-TRAP marker system. The dashed line indicates genetic distance at 0.11.

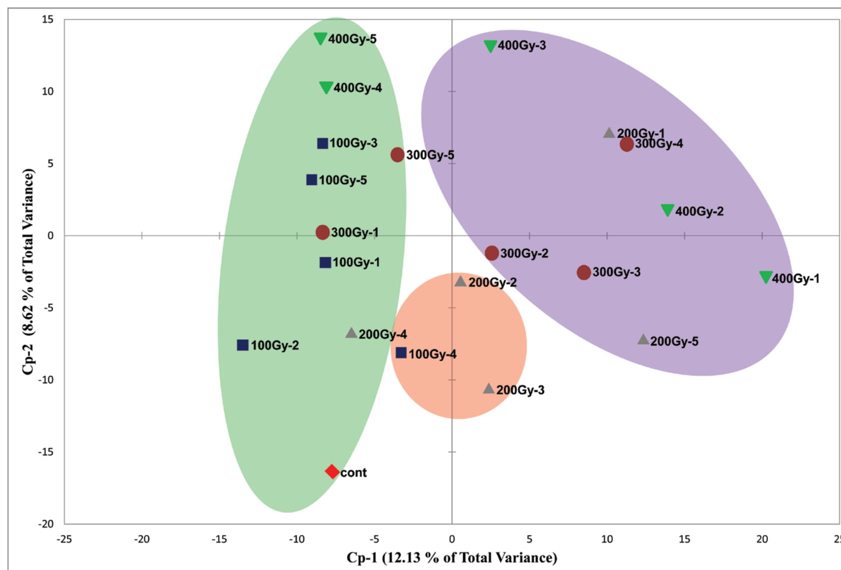

Figure 3. A two-dimensional principal component analysis of 21 ɣ-ray irradiated sorghum (*Sorghum bicolor*) plants using TE-TRAP marker system.

The genetic distance, differentiation, and AMOVA results were estimated to support the results of the phylogenetic tree and PCA for each marker system (Table 5). The genetic distance ranges of TE-TRAP, TD, and AFLP were 0.211-0.261, 0.137-0.173, and 0.124-0.209, respectively. The highest genetic distance and differentiation between the different groups were observed using the TE-TRAP marker system, whereas the lowest distance and differentiation was observed with the TD marker system. AMOVA conducted among and within each of the four treatment groups (for all five duplications of each dose) revealed that most of the molecular variation in the TE-TRAP marker system explained the most total variation with the among population estimated variance 21.48 and the within population estimated variance 212.61 (Table 6).

Table 5. Analysis of pairwise F_{ST} values and genetic distances from three different marker systems based on 21 ɤ-ray irradiated sorghum (*Sorghum bicolor*) individuals.

	Estimated variance	Percent variation
TE-TRAP		
Among treatment	21.48	9.2
Within treatment	212.61	90.8
Total	234.09	100.0
TD		
Among treatment	1.77	7.9
Within treatment	20.67	92.1
Total	22.44	100.0
AFLP		
Among treatment	5.32	10.2
Within treatment	46.74	89.8
Total	52.06	100.0

Genetic distances are shown above diagonal. F_{ST} values below diagonal. Negative F_{ST} values were converted to zero.

Table 6. Analysis of molecular variance (AMOVA) from three different marker systems on 21 ɤ-ray irradiated sorghum plants.

	Estimated variance	Percent variation
TE-TRAP		
Among treatment	21.48	9.2
Within treatment	212.61	90.8
Total	234.09	100.0
TD		
Among treatment	1.77	7.9
Within treatment	20.67	92.1
Total	22.44	100.0
AFLP		
Among treatment	5.32	10.2
Within treatment	46.74	89.8
Total	52.06	100.0

Association analysis between markers and doses

To study the association between the different markers and the ɤ-ray doses, 189 individuals (nine accessions with five individuals of each ɤ-ray dose and each original cultivars), which were bred after ɤ-ray irradiation, were tested using the TE-TRAP marker

system. The TE-TRAP marker system generated 1452 markers in the whole population (Figure 4) that were scored using a binary code (1 or 0 for presence or absence, respectively).

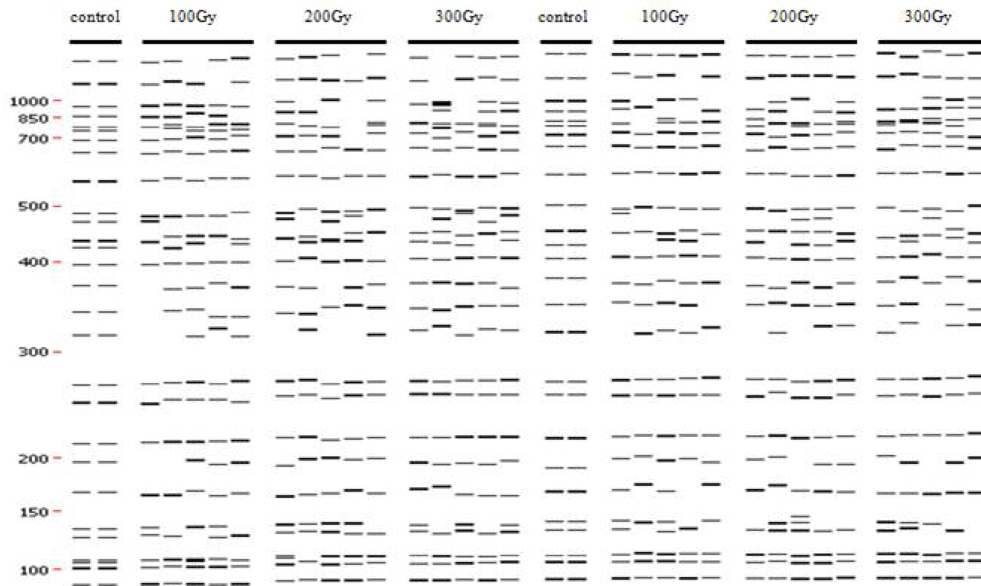

Figure 4. Resulting PCR product pattern of TE-TRAP system with Sb_PIF_2 + Sa4 primer combination for DINE-A-MITE cultivar. Two controls were used for validation of ample reproducibility.

The associations between the markers and ɤ-ray dose were explored using GLM and MLM + K methods. Two significant (P ≤ 0.005) associations between the marker and the irradiation dose were identified based on the average of the GLM and MLM + K methods (Table 7). The two markers were Sb_PIF_1a + sa4_15 and Sb_PIF_4 + sa4_35.

Table 7. Results of association analyses between 1452 TE-TRAP markers and different ɤ-ray doses (Control, 100, 200, and 300 Gy), using two statistical approaches (GLM and K+MLM).

TE-TRAP marker	GLM		K + MLM		Average P
	R^2 (%)	P	R^2 (%)	P	
Sb_PIF_1a + sa4_15	17.2	***	14.6	***	***
Sb_PIF_4 + sa4_35	12.4	***	12	**	***

***P ≤ 0.005, **P ≤ 0.01.

DISCUSSION

Sorghum is an important and promising global cereal crop. However, research on sorghum has fallen behind other cereal crops in terms of utilizing molecular biology tools to unwind the complexity of its genome and improve genetic capacity. The use of mutation breeding techniques provides a promising future for sorghum crop improvement. Accordingly, transposon-based marker systems have been developed to select for mutations.

Conventional TRAP marker arrays are a relatively new PCR-based marker technique (Hu and Vick, 2003). Application of the TRAP marker system to lettuce (Hu et al., 2005), sugarcane (Alwala et al., 2006), sunflowers (Yue et al., 2009), faba beans (Kwon et al., 2010), and chickpeas (Kumar et al., 2014) has suggested that TRAP is suitable for assessing genetic diversity, polymorphisms, and genetic relationships. TRAP uses a fixed primer designed to target expressed sequence tag sequences of candidate genes involved in controlling phenotypes, and amplifies fragments associated with a phenotype. This has been documented by Miklas et al. (2006) for a disease resistance trait in common beans, by Alwala et al. (2006) for sugar content in sugarcane, and by Yue et al. (2010) for the ray flower color in sunflowers.

In this study, TE-TRAP, TD, and AFLP marker systems were applied in 21 individuals to compare the genetic diversity, polymorphism, and genetic differentiation to confirm the efficiency and value of TE-TRAP. For the profile of each marker system, we identified that the TE-TRAP marker system generated more fragments than the other marker systems. TD amplified 223 fragments (avg. 14 amplicons), AFLP amplified 555 fragments (avg. 26 amplicons), and TE-TRAP generated a total of 1133 fragments (avg. 37 amplicons) per primer combination. The percent polymorphic fragments in the TD and AFLP marker systems were 50.8 and 58.4%, respectively, whereas it reached 63.0% in the TE-TRAP system. As shown in Table 4, the percent polymorphic amplicons per primer combination was large, ranging from 43.2 (Zm_PIF_1 + sa12) to 85.4% (Sb_PIF_3 + sa12) in the TE-TRAP marker system. The average PIC value of the TE-TRAP markers (0.172) was higher than for TD (0.122) or AFLP (0.132). The GD values for the TD and AFLP marker systems were 0.148 and 0.159, respectively, whereas the value for the TE-TRAP marker system was higher at 0.214. Hence, we confirmed the efficiency and applicability of the TE-TRAP marker system, which produced more polymorphic fragments and higher diversity. These results also show that the TE-TRAP marker system is applicable to ɤ-ray irradiated sorghum fingerprinting and that the system reveals polymorphisms with high efficiency. Although the TD and AFLP marker systems have high-throughput, both are more labor-intensive and less cost-efficient methods. Therefore, TE-TRAP is an attractive alternative that is easy to perform (not requiring endonuclease treatment), low-cost, and has the high-throughput associated with the use of ɤ-ray irradiation.

Notably, we compared the results of a conventional TRAP system and the TE-TRAP system developed in this study, to confirm differences between natural diversity and individual mutant diversity. One previously published study using a TRAP marker system on faba beans (Kwon et al., 2010) estimated the diversity value of faba entries as 0.091, whereas the diversity based on the TE-TRAP marker system for ɤ-ray irradiated sorghum was estimated at 0.214. These results suggest that mutation breeding increases genetic diversity, although it may have positive or negative impacts. In any event, the TE-TRAP marker system is suitable for identifying the level of diversity in mutation breeding plants.

In this study, the clustering patterns of the phylogenetic tree and the PCA based on genetic distance indicated a distinctive grouping of genotypes based on the TE-TRAP marker system. Three main groups were identified using the TE-TRAP marker system. However, the groups did not appear to be formed based on the ɤ-ray doses. It has been shown that physical mutageneses, such as ion-beam, x-ray, ɤ-ray, sometimes cause random variance (Kurowska et al., 2012; Kang et al., 2013). In contrast to the TE-TRAP, the phylogenetic tree and PCA based on the TD or AFLP systems did not result in any suitable groupings. We therefore suggest that the TE-TRAP marker system developed in this study allows for more accurate interpretation of the relationship between and among ɤ-ray irradiated sorghum individuals than the other

marker systems. In addition, our results indicate that TE-TRAP markers are useful for differentiating between irradiated sorghum individuals at each of the various ɤ-ray doses. The TE-TRAP array was the most sensitive and the most distinguishable marker system for ɤ-ray irradiation among the three marker systems. Tables 5 and 6 show the genetic distances and AMOVA results that were supported by grouping the data. The genetic distance and estimated variance of the TE-TRAP marker system were higher than those of the TD and AFLP systems. The highest genetic differentiation found in the TE-TRAP array supported our results that this system could identify the maximum genetic diversity among the marker systems and was the best marker system for ɤ-irradiated material.

In Table 5, negative F_{ST} values were converted to zero. These negative F_{ST} values indicate that individuals from different doses are genetically more closely related than individuals that received the same dose. As mentioned above, mutations induced by ɤ-rays appear to be random with respect to irradiation dose. By genotyping 189 ɤ-ray irradiated sorghum whole M_1 individuals with a TE-TRAP marker array, we conducted an analysis of associations between markers and the ɤ-ray doses. We found that the ɤ-ray dose was significantly associated with two markers ($P \leq 0.005$): Sb_PIF_1a + sa4_15 and Sb_PIF_4 + sa4_35. This result suggests that TE-TRAP may be used for the analysis of ɤ-ray irradiated sorghum and also as candidate genes to select initial mutations.

To date, the genomes of more than 90 plant species have been sequenced, these include 35 food crops. The number of whole plant genomes that are being sequenced is increasing exponentially and our TE-TRAP marker technique can be applied to more plant species to help select initial group mutations. This marker system is likely to be very useful for mutation breeding.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by grants from the Nuclear R&D Program by the Ministry of Science, ICT and Future Planning (MSIP: #2012M2A2A6003), and the research program of KAERI, Republic of Korea.

REFERENCES

Alwala S, Suman A, Arro JA, Veremis JC, et al. (2006). Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. *Crop Sci.* 46: 448-455. http://dx.doi.org/10.2135/cropsci2005.0274

Bennetzen JL, SanMiguel P, Chen M, Tikhonov A, et al. (1998). Grass genomes. *Proc. Natl. Acad. Sci. USA* 95: 1975-1978. http://dx.doi.org/10.1073/pnas.95.5.1975

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, et al. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635. http://dx.doi.org/10.1093/bioinformatics/btm308

Casa AM, Brouwer C, Nagel A, Wang L, et al. (2000). The MITE family heartbreaker (Hbr): molecular markers in maize. *Proc. Natl. Acad. Sci. USA* 97: 10083-10089. http://dx.doi.org/10.1073/pnas.97.18.10083

Doggett H (1988). Sorghum. 2nd ed. Longman, London; Published in the United states, Wiley, New York.

Doyle JJ and Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19: 11-15.

Feschotte C and Pritham EJ (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41: 331-368. http://dx.doi.org/10.1146/annurev.genet.40.110405.090448

Feschotte C, Jiang N and Wessler SR (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.* 3: 329-341. http://dx.doi.org/10.1038/nrg793

Hara Y, Abe T, Sakamoto K, Miyazawa Y, et al. (2003). Effects of heavy-ion beam irradiation in rose (*Rosa* hybrid cv. 'Bridal Fantasy'). *RIKEN Accel. Prog. Rep.* 36: 135.

Hu J and Vick BA (2003). Target region amplification polymorphism: a novel marker technique for plant genotyping. *Plant Mol. Biol. Report.* 21: 289-294. http://dx.doi.org/10.1007/BF02772804

Hu J, Ochoa OE, Truco MJ and Vick BA (2005). Application of the TRAP technique to lettuce (*Lactuca sativa* L.) genotyping. *Euphytica* 144: 225-235. http://dx.doi.org/10.1007/s10681-005-6431-1

Kang EJ, Lee YM, Sung SY, Ha BK, et al. (2013). Analysis of the genetic relationship of gamma-irradiated *in vitro* mutants derived from standard-type *Chrysanthemum* cv. Migok. *Hortic. Environ. Biotechnol.* 54: 76-81. http://dx.doi.org/10.1007/s13580-013-0124-9

Kikuchi K, Terauchi K, Wada M and Hirano HY (2003). The plant MITE mPing is mobilized in anther culture. *Nature* 421: 167-170. http://dx.doi.org/10.1038/nature01218

Kumar Y, Kwon SJ, Coyne CJ, Hu J, et al. (2014). Target region amplification polymorphism (TRAP) for assessing genetic diversity and marker-trait associations in chickpea (*Cicer arietinum* L.) germplasm. *Genet. Resour. Crop Evol.* 61: 965-977. http://dx.doi.org/10.1007/s10722-014-0089-2

Kunze R, Saedler H and Lönnig W (1997). Plant transposable elements. In: Advances in botanical research (Callow JA, ed.). Academic Press, Massachusetts, 27: 331-470.

Kurowska M, Labocha-Pawłowska A, Gnizda D, Maluszynski M, et al. (2012). Molecular analysis of point mutations in a barley genome exposed to MNU and gamma rays. *Mutat. Res.* 738-739: 52-70. http://dx.doi.org/10.1016/j.mrfmmm.2012.08.008

Kwon SJ, Hu J and Coyne CJ (2010). Genetic diversity and relationship among faba bean (*Vicia faba* L.) germplasm entries as revealed by TRAP markers. *Plant Genet. Resour.* 8: 204-213. http://dx.doi.org/10.1017/S1479262110000201

Lee JK, Kwon SJ, Park KC and Kim NS (2005). Isaac-CACTA transposons: new genetic markers in maize and sorghum. *Genome* 48: 455-460. http://dx.doi.org/10.1139/g05-013

Li G and Quiros CF (2001). Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica. Theor. Appl. Genet.* 103: 455-461. http://dx.doi.org/10.1007/s001220100570

Lombardi GM, Nunes JA, Parrella RA, Teixeira DH, et al. (2015). Path analysis of agro-industrial traits in sweet sorghum. *Genet. Mol. Res.* 14: 16392-16402. http://dx.doi.org/10.4238/2015.December.9.8

Liu K and Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129. http://dx.doi.org/10.1093/bioinformatics/bti282

McClintock B (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA* 36: 344-355. http://dx.doi.org/10.1073/pnas.36.6.344

Miklas PN, Hu J, Grünwald NJ and Larsen KM (2006). Potential application of TRAP (targeted region amplified polymorphism) markers for mapping and tagging disease resistance traits in common bean. *Crop Sci.* 46: 910-916. http://dx.doi.org/10.2135/cropsci2005.08-0242

Mueller UG and Wolfenbarger LL (1999). AFLP genotyping and fingerprinting. *Trends Ecol. Evol.* 14: 389-394. http://dx.doi.org/10.1016/S0169-5347(99)01659-6

Nei M (1972). Genetic distance between populations. *Am. Nat.* 106: 283-292. http://dx.doi.org/10.1086/282771

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551-556. http://dx.doi.org/10.1038/nature07723

Peakall R and Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295. http://dx.doi.org/10.1111/j.1471-8286.2005.01155.x

Rooney LW and Waniska RD (2000). Sorghum food and industrial utilization. In: Sorghum: origin, history, technology, and production (Smith CW and Frederiksen RA, eds.). Wiley, New York, 689-729.

Santiago N, Herráiz C, Goñi JR, Messeguer X, et al. (2002). Genome-wide analysis of the Emigrant family of MITEs of *Arabidopsis thaliana. Mol. Biol. Evol.* 19: 2285-2293. http://dx.doi.org/10.1093/oxfordjournals.molbev.a004052

Schneider S, Roessli D and Excoffier L (2000). Arlequin ver. 2.000. A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sneath PHA and Sokal RR (1973). Numerical taxonomy: the principles and practice of numerical classification. Freeman press, San Francisco.

Song HS and Kang SY (2003). Application of natural variation and induced mutation in breeding and functional genomics:

papers for international symposium; current status and future of plant mutation breeding. *Kor. J. Breed. Sci.* 35: 24-34.

Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599. http://dx.doi.org/10.1093/molbev/msm092

Vos P, Hogers R, Bleeker M, Reijans M, et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23: 4407-4414. http://dx.doi.org/10.1093/nar/23.21.4407

Wessler SR (2006). Transposable elements and the evolution of eukaryotic genomes. *Proc. Natl. Acad. Sci. USA* 103: 17600-17601. http://dx.doi.org/10.1073/pnas.0607612103

Yang G (2013). MITE Digger, an efficient and accurate algorithm for genome wide discovery of miniature inverted repeat transposable elements. *BMC Bioinformatics* 14: 186. http://dx.doi.org/10.1186/1471-2105-14-186

Yu SB, Xu WJ, Vijayakumar CH, Ali J, et al. (2003). Molecular diversity and multilocus organization of the parental lines used in the International Rice Molecular Breeding Program. *Theor. Appl. Genet.* 108: 131-140. http://dx.doi.org/10.1007/s00122-003-1400-3

Yue B, Cai X, Vick BA and Hu J (2009). Genetic diversity and relationships among 177 public sunflower inbred lines assessed by TRAP markers. *Crop Sci.* 49: 1242-1249. http://dx.doi.org/10.2135/cropsci2008.08.0471

Yue B, Vick BA, Cai X and Hu J (2010). Genetic mapping for the *Rf1* (fertility restoration) gene in sunflower (*Helianthus annuus* L.) by SSR and TRAP markers. *Plant Breed.* 129: 24-28. http://dx.doi.org/10.1111/j.1439-0523.2009.01661.x

Zhang X and Wessler SR (2004). Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. *Proc. Natl. Acad. Sci. USA* 101: 5589-5594. http://dx.doi.org/10.1073/pnas.0401243101