GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Characterization of Ty1/copia-like retrotransposon families from pigeonpea genome

P.G. Patil[1,2], M. Byregowda[1], I.O. Agbagwa[3] and H.E. Shashidhar[1]

[1]Department of Plant Biotechnology, University of Agricultural Sciences, GKVK, Karnataka, India
[2]Indian Institute of Pulses Research, Uttar Pradesh, India
[3]Department of Plant Science & Biotechnology, Faculty of Biological Sciences, University of Port Harcourt, Rivers State, Nigeria

Corresponding author: I.O. Agbagwa
E-mail: ikechukwu.agbagwa@uniport.edu.ng

Genet. Mol. Res. 14 (2): 5812-5822 (2015)
Received November 12, 2014
Accepted April 17, 2015
Published May 29, 2015
DOI http://dx.doi.org/10.4238/2015.May.29.13

ABSTRACT. Retrotransposons contribute significantly to the size, organization, and genetic diversity of their host genomes. To characterize novel retrotransposon families in pigeonpea and develop retrotransposon-based sequence-specific amplification polymorphic markers, *in silico* homology sequence search was carried out against the whole genome shotgun sequence of pigeonpea variety Asha (ICPL87119). For homology searching, 5 copia-like retro elements belonging to soybean, common bean, mungbean, chickpea, and field pea were used as query sequences. Contigs with at least 80% query coverage and >70% similarity were searched for retroelements using the long terminal repeat finder. A total of 28 copia-like retroelements were identified using this method. Multiple sequence alignment for the reverse transcriptase domain indicated conserved reverse transcriptase domains in all 28 elements compared with other reported elements. Phylogenetic analysis based on reverse transcriptase domains revealed

11 families. The copy number per family ranged from 1 (for B, J, and K family) to 8 (I). The sequence-specific amplification polymorphic marker-based insertion site profiling for one of the retrotransposon families (G) confirmed multiple insertions of this element across the pigeonpea genome. This study showed that our *in silico* homology search strategy was efficient for identifying and characterizing the Ty1/copia-like retrotransposon. The results of this study are useful for developing retrotransposon-based sequence-specific amplification polymorphic markers for pigeonpea crop improvement.

Key words: Sequence-specific amplification polymorphic markers; Long terminal repeat retrotransposons; Pigeonpea diversity

INTRODUCTION

Recent advances in sequencing of the pigeonpea whole genome have resulted in a tremendous increase in genomic resources for this crop (Singh et al., 2012; Varshney et al., 2012). Sequencing revealed a large number of transposable elements that have contributed to shaping of the pigeonpea genome. These elements are widely distributed in the plant genomes into 2 broad classes: Class I elements that transpose as a "copy and paste" mechanism and class II elements that transpose via a "cut and paste" mechanism. Further, Class I includes 2 types of elements, including long-terminal repeat-retrotransposons (LTR-RT) and non-LTR retrotransposons. LTR-retrotransposons bear LTRs at their terminal ends. Based on the order of internal domains (gag and pol open reading frames encoding protease, endonuclease, reverse transcriptase, and RNase H), LTR-retrotransposons are further divided into the Ty1-copia and Ty3-gypsy groups (Wicker et al., 2007). In the copia group, an endonuclease domain is positioned 5' to the reverse transcriptase domain, whereas in the gypsy group, this domain is 3' to the reverse transcriptase domain. LTR-retrotransposons are responsible for the vast differences in genomes size and genome arrangements in various plant species (Bennetzen, 2000).

Singh et al. (2012) reported that the total size of repeat elements (RE) in pigeonpea is 326.67 Mb, which was 63.95% of the 511 Mb available genome sequence. This is higher than the 23.90% reported in grape (Velasco et al., 2007), 25.03% in cucumber (Huang et al., 2009), 34.79% in rice (IRGSP, 2005), and 53.17% in papaya (Ming et al., 2008). The value is similar to the 61.47% in soybean (Schmutz et al., 2010), 67% in apple (Velasco et al., 2010), and 62% in sorghum (Paterson et al., 2009), but lower than the 84.20% in maize (Schnable et al., 2009). Most REs (92.8%) are of interspersed type, and thus are included in Class I (retrotransposons), Class II (DNA transposons), or unclassified transposable element transposons. Annotation of transposable elements in the genome sequence revealed a large number of retrotransposons compared to transposons. Retrotransposon insertions are irreversible, high in copy numbers, well-distributed throughout the genome, and changes remain relatively fixed, making them suitable candidates for characterizing and developing molecular tools for crop improvement programs.

In previous studies, various strategies have been applied to identify retrotransposon sequences in un-sequenced plant genomes. Pearce et al. (1999) developed a polymerase chain reaction (PCR)-based method to identify experimentally Ty1/copia-LTR sequences in higher plants, and isolated novel LTR sequences from pea, broad bean, and Norway spruce. However, the availability and increase in sequencing data for various crops may make *in silico* homolo-

gy searching very efficient for identifying retrotransposons. Only one Panzee a Ty1/copia-like element was described prior to pigeonpea whole genome sequencing (Lall et al., 2002). Based on the availability of the pigeonpea genome sequence, we selected few well-characterized copia-like elements belonging to different legumes as a query set. Computer-based sequence similarity searching was performed to identify homologous LTR-RT elements. Supported by phylogenic analysis, we distinguished retrotransposon families. Their phylogenic relationships with other reference plant retrotransposons were studied. Based on LTR divergence, insertion dating of full-length copies belonging to 7 pigeonpea retrotransposon families was estimated in order to examine the dynamics of these families. Finally, the sequence-specific amplification polymorphism (SSAP) profile of one retrotransposon family was evaluated in 30 pigeonpea genotypes.

MATERIAL AND METHODS

Plant material and DNA extraction

Thirty pigeonpea genotypes were used in this study to develop the SSAP profile (Table 1). The seeds of these genotypes were collected from the Indian Institute of Pulses Research, Kanpur, and All Indian Coordinated Research Project on pigeonpea, UAS, Bangalore. Genomic DNA was extracted from the young leaves of 15-day-old seedlings using a modified cetyl trimethyl ammonium bromide method suitable for legumes (Agbagwa et al., 2012).

Table 1. Details of 30 pigeonpea genotypes used for SSAP analysis.

S. No.	Pigeonpea genotypes	Species	Plant type
1	BRG 3	*Cajanus cajan*	NDT
2	ICP 8863	*Cajanus cajan*	NDT
3	ICP 7035	*Cajanus cajan*	DT
4	TTB 7	*Cajanus cajan*	NDT
5	ICP 15770	*Cajanus volublis*	DT
6	ICP 15853	*Rhynchosia rothi*	NDT
7	BDNP 3	*Rhynchosia rothi*	DT
8	ICP 817	*Rhynchosia bracteata*	DT
9	ICP 15890	*Rhynchosia rothi*	DT
10	BNG 1	*Cajanus scarabaeoides*	NDT
11	ICP 15815	*Rhynchosia bracteata*	DT
12	BDNP 4	*Cajanus albicans*	DT
13	HY 3C	*Cajanus cajan*	NDT
14	BRG 1	*Cajanus cajan*	DT
15	GRG 333	*Cajanus cajan*	NDT
16	GT 101	*Cajanus cajan*	NDT
17	BSMR 736	*Cajanus cajan*	NDT
18	WRP 1	*Cajanus cajan*	NDT
19	IPA 8F	*Cajanus cajan*	NDT
20	BRG 2	*Cajanus cajan*	DT
21	ICPL 87119	*Cajanus cajan*	NDT
22	JKM 189	*Cajanus cajan*	NDT
23	ICP 2376	*Cajanuscajan*	NDT
24	GRG 811	*Cajanus cajan*	NDT
25	TS3R	*Cajanus cajan*	NDT
26	ICP 15701	*Cajanus scarabaeoides*	DT
27	ICP 15667	*Cajanus platycarpus*	DT
28	ICPW 71	*Cajanus platycarpus*	DT
29	ICPW 61	*Cajanus platycarpus*	DT
30	ICP 15799	*Flemingia macrophylla*	NDT

DT-determinate plant type; NDT-non determinate plant type.

Data mining

In silico sequence homology search was conducted to characterize novel Ty1/copia elements. The representative *copia* like full/partial LTR-RT sequences from soybean (U96748.1), common bean (AY341443), mungbean (AY684686.1), chickpea (AJ535883.2), and field pea (X66399.1) were retrieved from the NCBI database and used as query sequences. Using the BLASTn algorithm, sequence searching was performed against pigeonpea whole genome shotgun sequences with >10X coverage (Singh et al., 2012). Contigs showing at least 80% coverage and >70% similarity with the query sequence were selected for sequence analysis.

Sequence analysis

The contigs with >70% similarity were searched for full-length retroelements using the online search tool LTR Finder ver. 1.02 (Xu and Wang, 2007). Sequences were edited using the Bio-Edit software (Hall, 1999). The intactness of identified retroelements was examined viz gag, pol, RH, RT, and LTR regions with perfect target site duplication. Amino acid sequences for reverse transcriptase gene were deduced from nucleic acid sequences using the ExPasy program (http://www.expasy.org/). The MEGA software version 4.0 (Tamura et al., 2007) was used to achieve multiple alignment of amino acid sequences by ClustalW and to draw the neighbor-joining tree using the Poisson distance model. Based on the RT domain, we grouped the retrotransposons into families when the amino acid identity with other members was ≥90% (Bowen and McDonald, 1999). Each member of the family was named as a *Cajanus cajan* retroelement (CcRT) with serial numbers.

Insertion dating for retroelements

The insertion times of LTR-retroelements were dated by aligning their 5' and 3' LTR sequences and identifying transition and transversion substitutions using the MEGA software package version 4.0 as described by San Miguel et al. (1998). The time for element insertion was calculated using the formula $T = K/2r$, where T = time, K = distance calculated using Kimura's 2-parameter model as implemented within the MEGA software package, and r = substitution rate. Kimura's 2-parameter model corrects for multiple hits (Kimura, 1980). The r-value for the substitution rate (1.3×10^{-8} per site per year) was used as described by Vitte and Bennetzen (2006). Finally, the ages of retroelements were represented as million years (My) since insertion.

SSAP primer design and PCR amplification

SSAP insertion patterns were produced largely following the protocol described by Waugh et al. (1997). Based on the ages of retrotransposon families, one of the retroelements (CcRT8) belonging to family G was selected. The sequence information of the 5' LTR region was used to design the following primer: 5'-GTGCTGGTGGCCTTTTCTCC-3'. Next, 0.5 µg genomic DNA was double-digested with 10 U *Eco*RI and 4 U *Mse*I enzymes at 37°C for 3 h. Digested DNA was ligated to a double-stranded adapter (5 pmol *Eco*RI and 10 pmol *Mse*I) in ligation buffer (10X *T4* DNA ligase buffer, 4 U *T4* DNA ligase enzyme) at 37°C for 16 h. These digested and ligated DNA templates were further diluted 1:7 and used for pre-ampli-

fication (primers complimentary to the *Eco*RI and *Mse*I adapters, with 1 selective nucleotide cytosine and adenine, respectively). PCR programs for pre-amplifications were: 1 cycle at 94°C for 3 min, 20 cycles at 94°C for 30 s, 56°C for 1 min, and 72°C for 1 min followed by 1 cycle at 72°C for 10 min. Pre-amplification products were further diluted 1:8 and used for selective amplification [selective *Eco*RI adapter-specific primers, with 3 selective nucleotides (CAG) used in combination with 1 retrotransposon-specific primer (CcRT8)]. PCR conditions for selective amplification were: 1 cycle at 94°C for 3 min, 11 cycles at 94°C for 30 s, and 65°C for 30 s with reduction by 1°C/cycle to 56°C. Amplification products were separated on 6% denaturing polyacrylamide gels and bands were resolved by silver staining (Bassam et al., 1991).

RESULTS

Identification and characterization of Ty1/copia-like retrotransposon families

The 5 partial/full-length unique *copia*-like LTR-RT sequences from related legumes used as query sequences identified 28 full-length homologous retrotransposons from the pigeonpea genome harboring an internal domain flanked by LTRs (Table 2). Fifteen sequences displayed perfect target duplication sites near the insertion regions. The domain order inside the *pol* gene of the amino acid sequence suggested that all LTR-RTs identified by *in silico* homology searching were Ty1/copia-like elements. Multiple sequence alignment for amino acid sequences located in the reverse transcriptase revealed domains similar to other well-characterized copia elements (Figure 1).

Table 2. List of homologous *copia* retrotransposons identified in the pigeonpea genome.

Query sequence used	Pigeonpea genome contigs with RT	RT name	Size of RT (bp)	Length 5-LTR/ 3-LTR (bp)	TSD 5'-3'	LTR identity (%)
Glycine max (U96748.1)	AFSP01009815.1	CcRT-1	4956	206/206	-	98.1
	AFSP01019198.1	CcRT-2	5087	275/257	-	90.2
Phaseolus vulgaris (AY341443)	AFSP01003916.1	CcRT-3	4623	350/312	AAAAG	85.4
	AFSP01018906.1	CcRT-4	4464	264/263	ATTTG	95.8
	AFSP01002032.1	CcRT-5	4471	265/266	AATAT	97.4
Vigna radiata (AY684686.1)	AFSP01000619.1	CcRT-6	2007	108/112	-	93.8
	AFSP01034138.1	CcRT-7	5245	496/499	-	90.2
	AFSP01018343.1	CcRT-8	4965	174/175	TTCT	98.9
Cicer arietinum (AJ535883.2)	AFSP01033887.1	CcRT-9	4660	114/114	GACCC	97.4
	AFSP01000667.1	CcRT-10	4723	114/114	ATGTT	97.4
	AFSP01021729.1	CcRT-11	3950	113/113	ACATT	96.5
	AFSP01004682.1	CcRT-12	4752	113/113	ACTTG	100
	AFSP01003435.1	CcRT-13	3190	114/114	CATGG	98.2
	AFSP01009554.1	CcRT-14	2197	159/159	-	95.6
	AFSP01010636.1	CcRT-15	5872	113/113	GCTTT	93.8
	AFSP01019021.1	CcRT-16	2226	115/115	AATAC	96.5
	AFSP01021784.1	CcRT-17	4712	183/184	AGATA	99.5
	AFSP01038345.1	CcRT-18	4771	184/184	GATTT	96.2
	AFSP01032996.1	CcRT-19	5527	184/184	CAACA	97.3
	AFSP01016111.1	CcRT-20	2474	155/178	-	85.4
	AFSP01040671.1	CcRT-21	4636	153/176	-	81.2
Pisum sativum (X66399.1)	AFSP01001855.1	CcRT-22	4138	176/174	-	98.3
	AFSP01017989.1	CcRT-23	4084	164/164	AGAA	96.3
	AFSP01029435.1	CcRT-24	4073	177/177	-	96.6
	AFSP01033458.1	CcRT-25	4071	175/175	-	98.3
	AFSP01012289.1	CcRT-26	4260	237/240	-	96.2
	AFSP01023864.1	CcRT-27	3177	165/165	-	95.8
	AFSP01030162.1	CcRT-28	4864	199/199	-	98.0

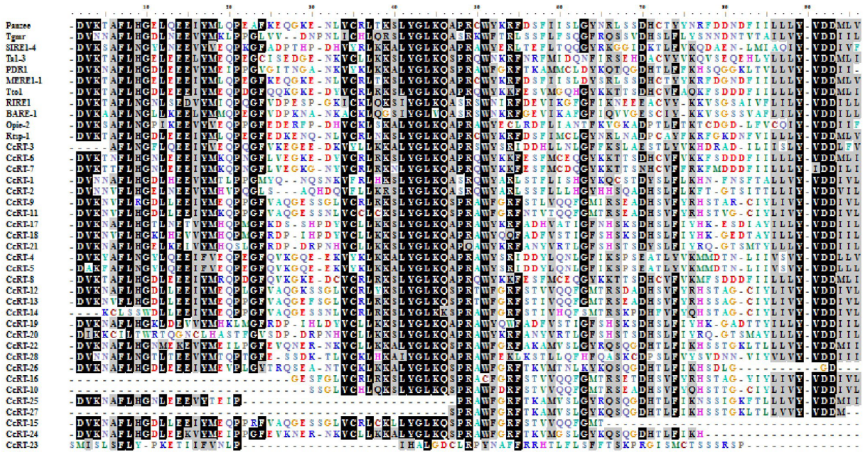

Figure 1. Multiple alignments for amino acids in RT domains for newly identified 28 CcRT pigeonpea sequence with other reported plant origin *copia* retrotransposons using ClustalW. The sequences used include *Cajanus cajan*, *Panzee* (accession No. AJ000893.1); *Glycine max copia*, *Tgmr* (U96748.1) and *SIRE1*-4 (AY205608.1); *Arabidopsis thaliana*, *Ta1*-3 (X13291); *Pisum sativum*, *PDR1* (X66399.1); *Medicago truncatula*, *MERE1*-1 (FJ544851.1); *Nicotiana tabacum*, *Tto1* (D83003.1), *Oryza australiensis*, *RIRE1* (D85597.1), *Hordeum vulgare*, *BARE*-1 (Z17327.1), *Zea mays*, *Opie*-2 (AF090446.1), and *Ipomoea batatas*, *Rtsp*-1 (AB162659.1).

The neighbor-joining tree clustered all 28 elements into 11 distinct families, which were also distinct from Panzee (Figure 2). The insertion number for each family ranged from 1 to 8 copies. The largest family, "I" included 8 elements, the "H" family 5 elements, the "G" family contained 3 elements, and 4 families, A, C, E, and F, each included 2 elements. Finally, 4 families, B, D, J, and K, were represented by a single element. The results of phylogenic assessment revealed the relationship between these families and other previously characterized retroelements in other crops [*Arabidopsis* (*Ta1*-3), alfalfa (*MERE1*-1), tobacco (*Tto1*), rice (*RIRE1*), barley (*BARE*-1), maize (*Opie*-2), and sweet potato (*Rtsp*-1)] (Figure 3).

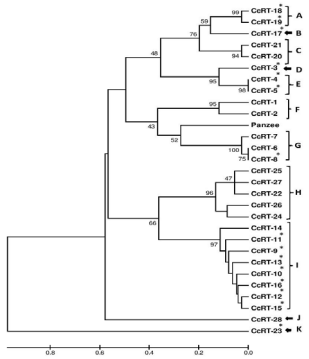

Figure 2. Classification of 28 CcRT pigeonpea retrotansposon sequences into 11 families. The neighbor-joining tree is based on the multiple sequence alignment of amino acid sequences between RT domains using the Poisson distance model. *Insertions displaying perfect target duplication sites. *Ty1*/*copia* retrotransposon of *Cajanus cajan*, *Panzee* (accession No. AJ000893.1) was added as reference.

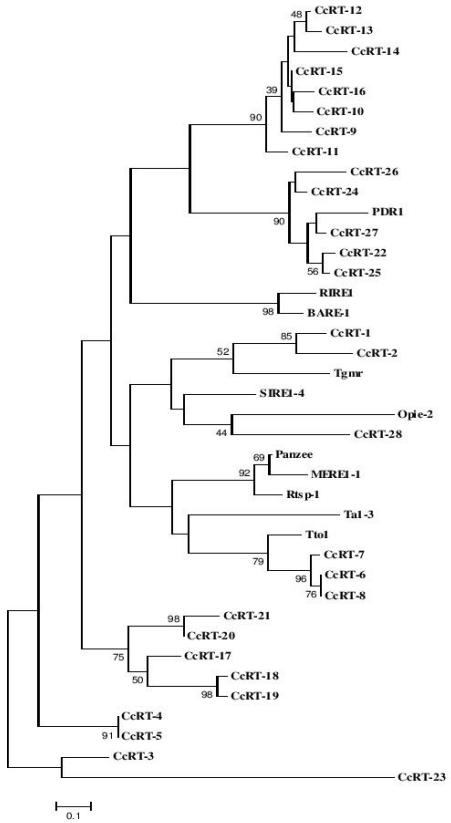

Figure 3. Phylogenetic analysis based on multiple sequence alignment of amino acid sequences for RT domain using the Poisson distance model for newly identified 28 CcRT pigeonpea sequence compared to other reported plant-origin *copia* retrotransposons. The sequences used included *Cajanus cajan*, *Panzee* (accession No. AJ000893.1); *Glycine max*, *Tgmr* (U96748.1) and *SIRE1*-4 (AY205608.1); *Arabidopsis thaliana*, *Ta1*-3 (X13291); *Pisum sativum*, *PDR1* (X66399.1); *Medicago truncatula*, *MERE1*-1 (FJ544851.1); *Nicotiana tabacum*, *Tto1* (D83003.1), *Oryza australiensis*, *RIRE1* (D85597.1), *Hordeum vulgare*, BARE-1 (Z17327.1), *Zea mays*, *Opie*-2 (AF090446.1), and *Ipomoea batatas*, *Rtsp*-1 (AB162659.1).

Age of retroelements

The results obtained for insertion dating for 7 families (A, B, D, E, G, I, and K) with at least 1 active element with perfect target site duplication revealed that the element copy age ranged from ≤1 to 14 million years (My) (0% divergence we represented as ≤ 1 My) (Table 3; Figure 4). Of the 18 elements, 9 (50%) were inserted within ≤ 1 My, 6 were inserted in the last 2.9 My, 2 were within 6.4 My, and 1 was within 14 My. Nine elements from 6 families showed maximum identity for 5' and 3' LTRs, indicating recent movement (0% divergence considered as insertion within the last ≤ 1 My in this study) in the pigeonpea genome. Insertion age patterns for a few families revealed discrete groups of copies inserted at the same date, such as in families I and G. This reflects different bursts of amplification for these families.

Table 3. Summary of parameters calculated to identify insertion dates for 7 families with retotransposons showing intact and perfect site duplications.

Pigeonpea contigs	RT	Family	K	SE	T	Age (My)
AFSP01038345.1	CcRT-18	A	0.168	0.131	6461538	6.4
AFSP01032996.1	CcRT-19	A	0	0	0	1
AFSP01021784.1	CcRT-17	B	0	0	0	1
AFSP01003916.1	CcRT-3	D	0.077	0.08	2961538	2.9
AFSP01018906.1	CcRT-4	E	0.077	0.08	2961538	2.9
AFSP01002032.1	CcRT-5	E	0	0	0	1
AFSP01000619.1	CcRT-6	G	0	0	0	1
AFSP01034138.1	CcRT-7	G	0.168	0.131	6461538	6.4
AFSP01018343.1	CcRT-8	G	0	0	0	1
AFSP01033887.1	CcRT-9	I	0	0	0	1
AFSP01000667.1	CcRT-10	I	0	0	0	1
AFSP01021729.1	CcRT-11	I	0.076	0.077	2923076	2.9
AFSP01004682.1	CcRT-12	I	0	0	0	1
AFSP01003435.1	CcRT-13	I	0.076	0.077	2923076	2.9
AFSP01009554.1	CcRT-14	I	0.364	0.201	14000000	14
AFSP01010636.1	CcRT-15	I	0.076	0.077	2923076	2.9
AFSP01019021.1	CcRT-16	I	0.077	0.08	2961538	2.9
AFSP01017989.1	CcRT-23	K	0	0	0	1
	Average				2476496	2.4

*Insertion times of LTR retrotransposons using LTR divergence method with a substitution rate of 1.3 x 10^{-8} substitutions per site per year. (RT = name of retrotransposon, K = Kimura's distance values, SE = standard error, T = time since insertion of element, My = million years, T = 0 considered as insertion within the last ≤1 My).

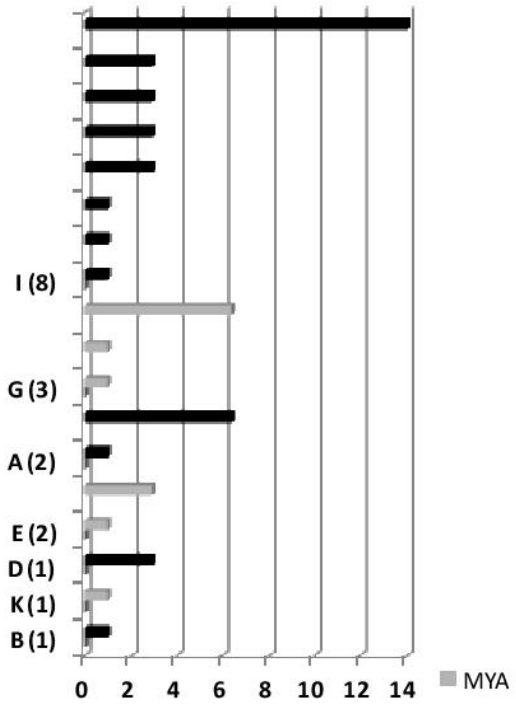

Figure 4. Distribution of insertion dates for full-length copies belonging to the 7 pigeonpea LTR retrotransposon families. The copy numbers of full-length elements for each family are shown in parenthesis.

Development of SSAP markers

SSAP profiling for the primer combination E-CAG/CcRT8 revealed numerous polymorphisms in the 30 genotypes examined (Figure 5). A total of 45 bands were scored and ranged in size from 100 to 600 base pairs, of which 42 bands were polymorphic (93.3%). These results show that based on insertion dating, elements with recent insertions can be used to develop SSAP markers.

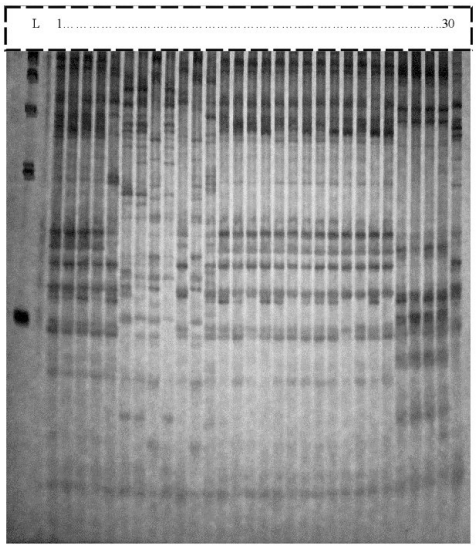

Figure 5. Amplification profiles for 30 pigeonpea genotypes using SSAP primer combination E-CAG/CcRT8 (*Lanes 1-30* represents pigeonpea genotypes as listed in Table 1). *Lane M* = 100-bp ladder.

DISCUSSION

In this study, we demonstrated the use of *in silico* homology searching for characterizing both homologous and novel retrotransposon families in the pigeonpea genome. Zhao et al. (2009) used a similar approach (*in silico* homology searching) to identify novel retrotransposons in the *Botrytis cinerea* genome. According to McClure et al. (1988), the reverse transcriptase domain showed the slowest relative rate of change among all retroelement proteins. Based on the amino acid conservation for the reverse transcriptase domain, we identified 11 Ty1/copia-like retrotransposon families. For phylogenetic analysis, we added Panzee, a Ty1/copia-like pigeonpea retrotransposon (Lall et al., 2002) to improve the comprehensiveness of our analysis. However, we found that none of the families identified in this study belonged to the pigeonpea retrotransposon Panzee. Therefore, to elucidate the relationship between these families and retroelements in other crops, further phylogenetic analysis was performed to include retroelements of *Arabidopsis* (*Ta1*-3), alfalfa (*MERE1*-1), tobacco (*Tto1*), rice (*RIRE1*), barley (*BARE*-1), maize (*Opie*-2), and sweet potato (*Rtsp*-1).

The *in silico* homology searching was very efficient for identifying and characterizing known homologous families in related crops. Homology searching can be used to detect numer-

ous homologous sequences as that of query sequence used; for example, RT elements belongs to family F (as soybean, *TGMR*), H (field pea, *PDR* 1), and I (chickpea, *cart157*). In addition, this method was helpful for characterizing novel families in the pigeonpea genome. Families K and G showed homology with the LTR-RTs of other crops such as maize (*Opie* 2) and tobacco (*Tto1*). Some novel families were identified, including families A and C; however, no families showed homology with Panzee, *RIRE1*, *SIRE1*-4, *MERE1*-1, *Rtsp*-1, *BARE*-1, and *Ta1*-3.

In order to develop SSAP markers, we calculated the insertion age for 7 families, as these families contained at least 1 active element with perfect target site duplication, and selected family G for the development of SSAP markers in pigeonpea. Because pigeonpea is widely grown in tropical areas, this element may contribute to insertional polymorphisms for creating wider variation in pigeonpea. SSAP profiling based on this family revealed a higher level of polymorphism (93.3%) across the 30 genotypes tested. Similarly, in pigeonpea *Panzee* retrotransposon-based SSAP markers, a previous study by Patil et al. (2012) revealed higher polymorphism (90.71%) across 21 genotypes. Our strategy enabled characterization of a few novel families of copia elements for the first time in pigeonpea. Finally, our method enabled selection and development of SSAP markers based on insertion dating.

REFERENCES

Agbagwa IO, Datta S, Patil PG, Singh P, et al. (2012). A protocol for high-quality genomic DNA extraction from legumes. *Genet. Mol. Res.* 11: 4632-4639.

Bassam BJ, Caetano-Anollés G and Gresshoff PM (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.* 196: 80-83.

Bennetzen JL (2000). Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.* 42: 251-269.

Bowen NJ and McDonald JF (1999). Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral like elements. *Genome Res.* 9: 924-935.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41: 95-98.

Huang S, Li R, Zhang Z, Li L, et al. (2009). The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 41: 1275-1281.

International Rice Genome Sequencing Project (2005). The map based sequence of the rice genome. *Nature* 436: 793-800.

Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.

Lall I, Maneesha and Upadhyaya KC (2002). *Panzee*, a *copia* like retrotransposon from the grain legume, pigeonpea (*Cajanus cajan* L.). *Mol. Genet. Genomics* 267: 271-280.

McClure MA, Johnson MS, Feng DF and Doolittle RF (1988). Sequence comparisons of retroviral proteins: Relative rates of change and general phylogeny. *Proc. Natl. Acad. Sci. U. S. A.* 85: 2469-2473.

Ming R, Hou S, Feng Y, Yu Q, et al. (2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* L.). *Nature* 452: 991-996.

Paterson AH, Bowers JE, Remy B, Inna D, et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457: 551-556.

Patil P, Datta S, Singh IP, Das A, et al. (2012). Phylogenetic analysis of pigeonpea (*Cajanus cajan*) genotypes using *Panzee* retrotransposon based SSAP markers. *Indian J. Agric. Sci.* 82: 1016-1021.

Pearce S, Stuart Rogers C, Knox M, Kumar A, et al. (1999). Rapid isolation of plant *Ty1-copia* group retrotransposon LTR sequences for molecular marker studies. *Plant J.* 19: 711-717.

San Miguel P, Gaut BS, Tikhonov A, Nakajima Y, et al. (1998). The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20: 43-45.

Schmutz J, Cannon SB, Schlueter J, Ma J, et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178-183.

Schnable PS, Doreen W, Fulton RS, Stein JC, et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112-1115.

Singh NK, Gupta DK, Jayaswal PK, Mahato AK, et al. (2012). The first draft of the pigeonpea genome sequence. *J. Plant Biochem. Biotechnol.* 21: 98-112.

Tamura K, Dudley J and Nei M (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.

Varshney RK, Chen W, Li Y, Bharti AK, et al. (2012). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30: 83-89.

Velasco R, Andrey Z, Michela T, Cartwright DA, et al. (2007). A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* 12: 313-326.

Velasco R, Zharkikh A, Jason A, Amit D, et al. (2010). The genome of the domesticated apple (*Malus* x *domestica* Borkh.). *Nat. Genet.* 42: 833-839.

Vitte C and Bennetzen JL (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* 103: 17638-17643.

Waugh R, McLean K, Flavell AJ, Pearce SR, et al. (1997). Genetic distribution of *BARE*-1 like retrotransposable elements in the barley genome revealed by sequence specific amplification polymorphisms (SSAP). *Mol. Gen. Genet.* 253: 687-694.

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

Xu Z and Wang H (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucl. Acids Res.* 35: 265-268.

Zhao M, Zhou JY, Li ZD, Song WW, et al. (2009). *Boty-II*, a novel LTR retrotransposon in *Botrytis cinerea* B05.10 revealed by genomic sequence. *Elec. J. Biotechnol*. 12: 1-9.

