Rani, A., Kumar, V., Shukla, S., Jha, P. and Rawal, R. (2016), *Seed Sci. & Technol.,* **44**, 357-369.
http://doi.org/10.15258/sst.2016.44.2.18

DNA barcoding of Indian soybean varieties as constructed through SSR markers

A. RANI, V. KUMAR*, S. SHUKLA, P. JHA AND R. RAWAL

ICAR-Indian Institute of Soybean Research, Indore, Madhya Pradesh, 452001, India
(E-mail: vineetksahni@yahoo.com)

(Accepted May 2016)

Abstract

102 Indian soybean varieties were surveyed using 10 SSR markers that were selected based upon high polymorphic information content (PIC) observed in the initial screening of 40 randomly selected genotypes using 58 SSR markers. The 10 selected primer pairs amplified 3-8 alleles in the 102 varieties. In total, 50 alleles with amplicon size ranging from 100 to 330 bp were observed with PIC value ranging from 0.4760 (primer pair Satt229) to 0.8123 (Sct_199). Once the amplicon profile of all the varieties was obtained, alleles were assigned a numerical number in the order of increasing size of amplicon. The numerical numbers were placed from left to right in alphabetical order of linkage group of the 10 SSR markers to construct a 10-digit barcode, which would serve as unique identification code for each of 102 soybean varieties released for commercial cultivation in India and would be useful in testing their genetic purity.

Introduction

Globally, India ranks fifth among seven major soybean-producing countries. To date, 102 soybean varieties have been developed and released for commercial cultivation in the country. These varieties have contributed immensely in raising soybean production from 0.14 mt ha^{-1} in 1970 to the highest level of 14.1 mt ha^{-1} in 2013. In the backdrop of the Protection of Plant Varieties and Farmers Regulatory Act (PPVFRA), identification of crop cultivars including soybean during breeding, production and processing is very important to protect plant breeders' rights. Identification of Indian soybean varieties using morphological descriptors alone is difficult due to their narrow genetic base. Most varieties of them have genetically similar parentage. Further, morphological traits/descriptors such as growth-type, days-to-flowering, flower colour, leaf shape, leaf colour, plant height, pod pubescence, stem pubescence, pod colour, days-to-maturity, seed size, seed coat colour, seed hilum colour and seed luster, conventionally used for identification of the cultivars, are limited and influenced by the environment (Khan *et al*., 2003).

* Author for correspondence

Moreover, for distinguishing varieties based upon morphological descriptors, a grow-out test is required, which is a time-consuming process. Isozymic electropherograms depicting variation in biochemical components of the seeds of the varieties are not useful for varietal identification due to their limited variability and influence of environmental factors on their *de novo* synthesis. Therefore, it is important to differentiate these varieties accurately through DNA markers in addition to morphological descriptors.

DNA fingerprinting techniques are the most reliable techniques to precisely, distinctly and rapidly distinguish morphologically-alike cultivars. According to the International Union for the Protection of New Varieties of Plants (UPOV), results based upon DNA markers constitute the legal basis for varietal protection. Guidelines for DNA profiling for identification of varieties adopted by UPOV has also been published (UPOV/ INF/17/1 2010). DNA fingerprinting techniques such as Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Simple Sequence Repeats (SSR), Inter Simple Sequence Repeat (ISSR) and Sequence Tagged Sites (STS) markers are useful in cultivar identification. SSR markers have been widely preferred for varietal identification due to their abundance, codominant inheritance, polymorphism, reproducibility and convenience in application through PCR. SSR fingerprinting for the identification of cultivars has been employed in several crops: potato (Novakova *et al*., 2010), rice (Zhu *et al*., 2012), wheat (Zhu *et al*., 2011), cotton (Ahemd *et al*., 2013), chickpea (Joshi *et al*., 2013), maize (Sudharani *et al*., 2014), and chrysanthemum (Zhang *et al*., 2014). Gao *et al.* (2009) assigned the molecular identity code to soybean varieties grown in the Heilongjiang province of China using simple sequence repeat markers. In Brazil, the dispute over soybean variety Conquista grown in 3 million ha and newly released variety. Pioneers which was suspected to be very similar to Conquista, could be resolved using application of molecular markers, when the results showed that the two cultivars represented the same genotype. Tantasawat *et al*. (2011) used SSR markers in identification of soybean varieties in Thailand. We differentiated six commercial varieties and six non-traded soybean varieties using nine SSR markers in a previous study (Kumar *et al*., 2015). In India, morphological descriptors could characterise only 20 soybean varieties, of the 102 soybean varieties released for cultivation in the country. In the present investigation, we subjected all the 102 soybean varieties to SSR marker analysis with high PIC value and assigned all of them a 10-digital molecular identity code, which is equivalent to the barcode given to the articles scanned by the infrared meter to ascertain their respective identity and value.

Materials and methods

Plant material

102 soybean varieties (table 1) were raised in the fields of the ICAR-Indian Institute of Soybean Research in Indore and Madhya Pradesh, India. DNA was extracted from the seeds of each variety following the method validated for DNA extraction from soybean seeds (Community Reference Laboratory for GM Food and Feed, 2007).

Table 1. Indian soybean varieties and their pedigree.

Variety	Pedigree
ADT-1(UGM-33)	Selection from "Hill variety"
Alankar	D 63-6094 × D61-4249
Ankur	SPS from composite of 22 crosses
Birsa soybean	Spontanous mutant of 'Sepaya Black'
Bragg	Jackson × D 49-2491
Co-1	Selection from EC398321
Co Soya-2	UGM21 × JS335
Co -3	UGM 69 × JS335
DS 228	JS335 × DS 181
Pusa97-12	Mutant of DS74
Davis	Roanoke×(Odgen×CNS) ×(Ralsoy×odgen)
Gujrat soybean 1 (J-231)	Selection from Punjab-1
Gujrat soybean 2 (J-202)	Selection from Geduld variety
Harit soy (Himso 1563)	Himso 1520 × Bragg
Hardee	D 49-772 × Improved Pelican
Indira soya 9	Secondary selection from JS 80-21
Improved Pelican	Tanloxi × PI60406
JS 2	Selection from Tehri Garhwal material.
JS20-34	JS98-63 × PK768
JS 71-05	Selection from Lee type exotic material
Gaurav (JS72-44)	D 60-9647 × EC 7034
Durga (JS 72-280)	EC14437 × Bragg
JS75-46	Improved Pelican × Semmes
JS76-205	Kalitur × Bragg
JS79-81	Bragg × Harsoy-Decid-uous
JS80-21	JS75-1 × PK 73-94
JS90-41	PS73-7 × Hark
JS93-05	Secondary selection from PS 73-22
JS95-60	Selection from PS 73-22
JS97-52	PK 327 × L 129
JS 335	JS78-77 × JS71 -5
JS2029	JS97-52 × JS95-56
Kalitur	Indigenous native variety
KHSb2	Manloxi × EC 39821
KB79 (Sneh)	Hardee × Monetta
Lee	S-100 × CNS
LSb1	Selection from MACS 330
MACS13	Hampton × EC 7034
MACS57	JS2 × Improved Pelican
MACS58	JS2 × Improved Pelican
MACS124	JS 2 × Improve Pelican
MACS450	Bragg × MACS 111
MAUS1 (Aarti)	Mutant from DS 87-14
MAUS2 (Pooja)	Selection from SH 84-14
MAUS61 (Pratikar)	JS71-1 × PK-73-94
MAUS61-2 (Pratishta)	JS80-21 X KB-60
MAUS71 (Samrudhi)	JS 71-5 × JS 87-38
Monetta	Exotic variety EC 2587
MAUS158	-
MAUS47	PS73-7 × Hark
MAUS81 (Shakti)	KB74 × JS335
MAUS 32 (Prasad)	Selection from JS 80-21
NRC2 (Ahilya-1)	Induced mutant of Bragg
NRC 37 (Ahilya 4)	Gaurav × Punjab 1
NRC7 (Ahilya-3)	Selection from S 69-96
NRC12 (Ahilya-2)	Induced mutant of Bragg
NRC86	RKS15×EC481309
PRS 1	Selection from germ-plasm
Pusa16	CNS × Lee

Table 1. *cont'd*

Table 1. *Continued.*

Variety	Pedigree	Variety	Pedigree
Pusa20	Bragg × L ee	RKS24	PK472 ×PK1024
Pusa24	Shelby × Bragg	RAUS5 (Pratap Soya 1)	Pusa16 × JS335
Pusa 37	Bragg × Java 16	SL4	EC 7965 × Bragg
Pusa 98-14	Bragg × DS 93-MM-39	Shilajeet	Selection from EC 9309
PK 262	UPSM97 × Hardee	SL96	Botato × JS 3
PK471	Hardee × Punjab1	SL 688	PK416 × SL317
Palam soya	-	Shivalik	Selection from segregating PK 73-55
PK 327	UPSM 82 × Semmes	SL295	PK416 × PK564
Pratap Soya 2	MACS 450 × Monetta	SL525	PK 416 × PK1023
Pusa22	Punjab 1 × Clark 63	Type 49	Selection from indigenous material
Punjab1	Selection from Nanking variety	TAMS 38	Monetta × PK472
Pusa40	8-3 × Lee	TAMS 98-21	Mutant of JS80-21
PK416	UPSM 534 × S 38	VLS1	Mutant of Bragg
PS1042	Bragg × PK 416	VLS2	Selection from VHC 856007
PS1024	PK 308 × PK317	VLS21	Selection from VHC 3055
PS 1225	PK515 × PK327	VLS47	Selection from KHSF-3-1-1
PS1029	PK 262 × PK317	VL S63	VLS2× (Bragg × VHC3022)
PK 308	T 31 × Hardee	VLS 65	Selection from local cultivar
PK472	Hardee × Punjab-1	VLS59	(Pb1 × VLS2) × EC361336
PK564	(UPSM 534 × Ankur) × Bragg		
PS1092	PK 327 × PK416		
PS1241	PK 1039 × PK327		
PS1347	PS 1024 × PK472		
RVS2001-4	JS93-01 × EC390981		

Selection and synthesis of primer pairs

Selection of SSR markers for surveying 102 soybean varieties was paramount for constructing barcodes for each variety. Initially, we screened 40 randomly selected soybean varieties using 58 SSR markers spanning 20 linkage groups. Ten SSR markers (Satt538, Satt577, Satt267, Satt146, Satt352, Sct_199, Satt541, Satt181, Satt229 and Satt009) showing polymorphic information content (PIC) values ≥ 0.476 were employed for surveying the 102 soybean varieties (table 2). The synthesis of the oligonucleotide sequences of 58 SSR primer pairs was outsourced to Sigma Aldrich, Bangalore. The sequences of SSR markers are from the list of soybean SSR loci mapped by Agricultural Research Services, United States Department of Agriculture, and are available at bldg6. arsusda.gov/cregan/soymap.htm.

Table 2. Primer sequences, linkage groups of different primers used for SSR profiling of 102 soybean varieties with their amplicon size.

S. No	SSR marker	LG	Sequences of forward and reverse primers	PIC Value	No. of alleles	Amplicon size (bp)
1	Satt538	A2	F 5′-GCAGGCTTATCTTAAGACAAGT-3′ R 5′-GGGGCGATAAACTAGAACAGGA3′	0.4940	3	110, 115, 125
2	Satt577	B2	F 5′-CAAGCTTAAGTCTTGGTCTTCTCT-3′ R 5′-GGCCTGACCCAAAACTAAGGGAAGTG-3′	0.6405	4	100, 110, 120, 130
3	Satt267	D1a	F 5′-CCGGTCTGACCTATTCTCAT-3′ R 5′-CACGGCGTATTTTTATTTTG-3′	0.6328	4	220, 230, 240, 250
4	Satt146	F	F 5′-AAGGGATCCCTCAACTGACTG-3′ R 5′-GTGGTGGTGGTGAAAACTATTAGAA-3′	0.6960	6	280, 290, 300, 310, 320, 330
5	Satt352	G	F 5′-GCGAATGTATTTTTGTTTCTCCATCAA-3′ R 5′-TGATAAGCCAAAAAATGGAAGCATAG-3′	0.7487	6	160, 170, 180, 185, 190, 200
6	Sct_199	G	F 5′-GCGACAATGGCTATTAGTAACAATCA-3′ R 5′-GCGATTTTCTATTTTCCTCACAGTG-3′	0.8123	7	190, 200, 205, 210, 220, 225, 240
7	Satt541	H	F 5′-GCGAATCCATCACACATAAA-3′ R 5′GCGGTACTCCCTCCAGAAAATAACC-3′	0.5821	5	150, 160, 165, 170, 190
8	Satt181	H	F 5′-TGGCTAGCAGATTGACA-3′ R 5′-GGAGCATAGCTGTTAGGA-3′	0.4913	3	180, 210, 220
9	Satt229	L	F 5′-TGGCAGCACACCTGCTAAGGGAATAAA-3′ R 5′-GCGAGGTGGTCTAAAATTATTACCTAT-3′	0.4760	4	180, 200, 205, 225
10	Satt009	N	F 5′-CCAACTTGAAATTACTAGAGAAA-3′ R 5′-CTTACTAGCGTATTAACCCTT-3′	0.8075	8	170, 180, 190, 200, 210, 220, 230, 250

PCR conditions and analysis

The PCR mixture contained 2 µl DNA (20 ngµl^{-1}), 1 µl PCR (10×) buffer, 1.1 µl $MgCl_2$ (25 mM), 0.1 µl dNTPs (25 mM), 0.4 µl each forward and reverse SSR primers (30 ng µl^{-1}), 0.068 µl Taq DNA polymerase (3 U µl^{-1}), and 4.932 µl distilled water. In the pre-cycle, DNA was denatured at 94°C for one minute followed by 30 cycles comprising denaturation at 94°C for one minute, primer annealing at 50°C for two minutes and primer elongation at 72°C for three minutes. Final elongation was at 72°C for 10 minutes. A Thermocycler model PTC100 was used for PCR, while PCR products were resolved on 3% metaphore gel. The PIC value of SSR markers was calculated as $PIC_i = 1-\sum^{n}_{j=1} p^2_{ij}$ where *i* denotes the SSR marker while p_{ij} is frequency of j^{th} allele. The frequency is the number of times a particular allele appeared in 102 varieties divided by total number of DNA bands generated in the whole of population.

Results

Detection of polymorphic alleles

SSR primer pairs represent loci and the DNA bands are alleles. Ten SSR primer pairs amplified 3-8 alleles, with amplicon size ranging from 100 to 330 bp (table 2). In total, 50 alleles with an average number of 5.0 alleles per locus were detected. Primers pairs with allele numbers greater than the average (5.0) were Satt009 (8), Sct_199 (7), Satt146 (6) and Satt352 (6). Three alleles were observed at Satt538 and Satt181; four alleles at Satt267, Satt577 and Satt229; five alleles at Satt541. The PIC value of these SSR primer pairs ranged from 0.4760 (Satt229) to 0.8123 (Sct_199). Amplicon profile generated by primer pair Sct_199, Satt577 and Satt229 for the 102 cultivars is shown in figure 1.

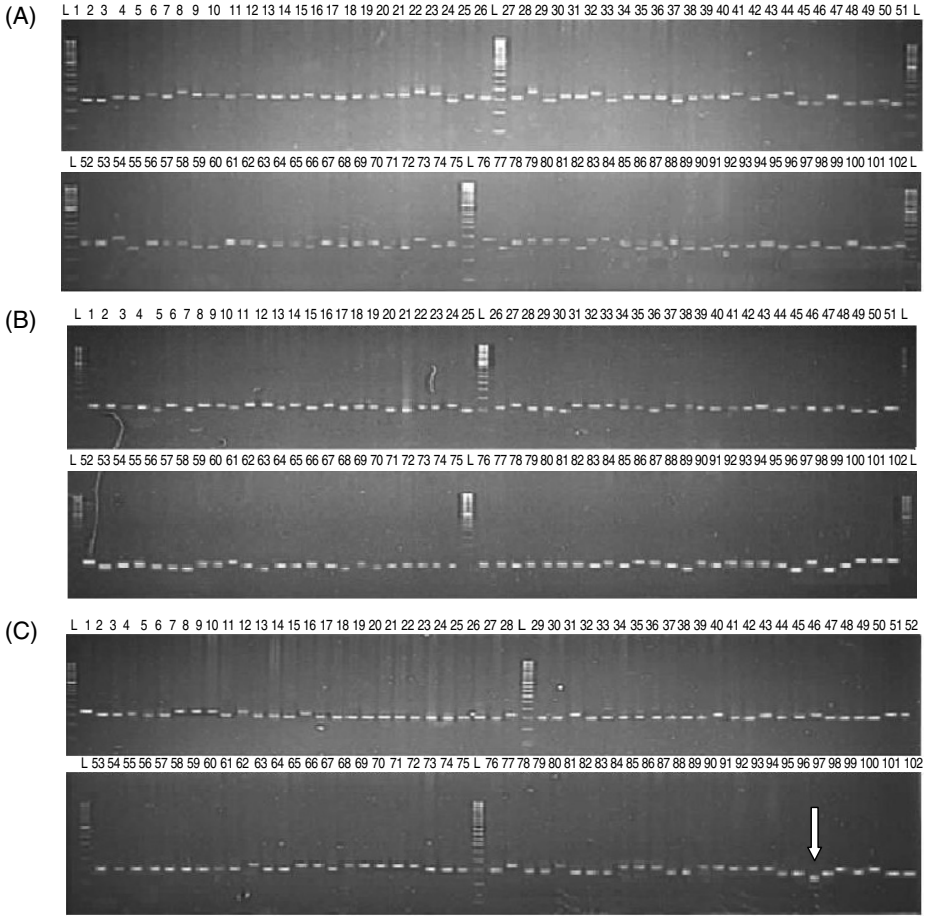

Figure 1. The amplicon profile of 102 Indian soybean cultivars generated through Sct_199 (**A**), Satt577 (**B**) and Satt229 (**C**), respectively. The 180 bp fragment shown by the arrow corresponds to a unique allele generated in variety JS20-34. Lanes 1 to 102 correspond to amplicons generated by 102 varieties, while lane L corresponds to the 50 bp DNA ladder.

Unique alleles

Among 50 alleles detected, four were unique alleles, generated in specific varieties (table 3). SSR marker Satt541 (linkage group H) generated two unique alleles, 165 bp fragment was amplified in variety Co-3 and a 190 bp amplicon in variety PK471 only. Satt229 (Linkage group L) amplified unique allele of 180 bp in variety JS20-34. Satt267 (Linkage group D1a) generated an exclusive allele of 250 bp in variety LSb1. Thus, unique alleles generated by specific primer pairs distinguished four varieties, Co3, PK471, JS20-34 and LSb1.

Table 3. Unique and rare alleles specific to varieties.

Unique alleles				Rare alleles			
Primer pair	Linkage group	Size (bp)	Variety	Primer pair	Linkage group	Size (bp)	Varieties
Satt541	H	165	Co-3	Sct_199	G	205	CO-3, Puss22, Co-1, Monetta
		190	PK471	Satt009	N	170	Gujarat soybean 2 , MACS81, LSb1, SL525, MACS13, JS72-44 ,VLS47, MAUS158, JS72-280
						200	Harit soy, Shilajeet, MAUS71, PS1029, RAUS 5, PS1024
						220	PRS1, Pusa24, VLS21, PS1225, Davis, PK564, VLS1, Pusa37
						250	CO-1,VLS63
Satt229	L	180	JS20-34	Satt352	G	180	MACS13, NRC86, JS20-34, Davis, JS72-44, SL96, NRC37, Monetta, JS335
				Satt146	F	280	Gujarat soybean 2, Birsa soybean 1, JS 2, Palam soya
						330	ADT1, SL295, PK308,SL525
				Satt577	B2	130	VLS 65, JS72-280
				Satt538	A2	115	Pusa16, Pusa37
Satt267	D1a	250	LSb1	Satt229	L	205	PRS1, PK564

Rare alleles

According to UPOV guidelines, the rare alleles i.e. alleles at a specific locus which appear with a frequency below an agreed threshold (commonly 5-10%), may be employed in cultivar identification (UPOV/INF/17/1 2010). In the present study, 11 rare alleles were detected, which appeared in two or more cultivars. The 205 bp allele at locus Sct_199 (linkage group G) appeared in four varieties (Co-3, Pusa22, Co-1 and Monetta); the 170 bp amplicon generated at Satt009 (LGp N) in nine varieties (Gujrat soya 2, MACS81, LSb1, SL525, MACS13, JS72-44, VLS47, MAUS158 and JS72-280). The same primer pair

generated amplicons of 200 bp in six varieties (Harit soya, Shilajeet, MAUS71, PS1029, RAUS5 and PS1024), 220 bp in eight varieties (PRS1, Pusa24, VLS21, PS1225, Davis, PK564, VLS1 and Pusa37) and 250 bp in two varieties (Co-1 and VLS63). At locus Satt352 (linkage group G), a 180 bp allele appeared in nine varieties (MACS13, NRC86, JS20-34, Davis, JS72-44, SL96, NRC37, Monetta and JS335). Satt146 (linkage group F) generated a 280 bp allele in four varieties (Gujarat soybean 2, Birsa soybean 1, JS2 and Palam Soya). The same primer pair generated a 330 bp amplicon in four varieties (ADT1, SL295, PK308 and SL525). At locus Satt577 (linkage group B2), an allele of 130 bp was detected in varieties VLS65 and JS72-280. Satt538 (linkage group A2) generated an amplicon of 115 bp allele in varieties Pusa16 and Pusa37; while Satt229 (linkage group L) amplified an allele of 205 bp in PRS1 and PK564. Thus, the afore-mentioned 11 alleles could identify a set of 2-9 varieties.

Construction of DNA barcode

Once all the varieties were surveyed using the 10 SSR primer pairs, alleles generated at each locus were assigned a number (1, 2, 3…) in the order of increasing size of the amplicons generated. For instance, the three alleles 110, 115 and 125 bp generated at locus Satt538 in our study were assigned 1, 2, and 3, respectively. Based upon the amplicon profile generated by surveying 102 varieties using 10 primer pairs, a 10-digit DNA barcode for each variety was constructed (table 4). For this purpose, the assigned number of each allele for 10 primer pairs was placed from left to right in alphabetical order of their linkage groups. Digits from left to right corresponded to the allele at loci Satt538, Satt577, Satt267, Satt146, Satt352, Sct_199, Satt541, Satt181, Satt229 and Satt009. For instance, barcode for the variety MACS13 was 1314372341, which from left to right signified 1st, 3rd, 1st, 4th, 3rd, 7th, 2nd, 3rd, 4th and 1st allele of Satt538, Satt577, Satt267, Satt146, Satt352, Sct_199, Satt541, Satt181, Satt229 and Satt009, respectively. This procedure of constructing barcodes was in line with the molecular identity code allocated to Chinese varieties from Heilongjiang Province by Gao e*t al*. (2009).

Discussion

Varieties that have been developed through selection from the same genotype or with common parentage shared the same digits at several positions in the 10-digital barcode (tables 1 and 4). The barcode of two genotypes itself would reflect how closely the varieties are related to each other. Two popular soybean varieties JS93-05 and JS95-60, selected from same genotype PS73-22, were assigned barcodes of 1334354227 and 1334454127, respectively. These two varieties can be identified by the allelic differences at two loci, Satt352 (5th digit) and Satt181 (8th digit). Satt352 generated alleles of 180 and 185 bp, while Satt181 generated amplicons of 210 and 180 bp in JS93-05 and JS95-60, respectively. PK471 (1214425122) and PK472 (1224644145) derived from the same cross Hardee × Punjab1 exhibited same alleles at four loci. Similarly, MAUS47 (3123451122) and JS90-41 (1314444247) with same parentage (PS73-7 and Hark) can be distinctly distinguished from each other by their barcodes.

Table 4. DNA barcodes of 102 soybean varieties as constructed using 10 SSR markers.

Variety	Barcode	Variety	Barcode	Variety	Barcode
ADT-1	3226651143	KHSb2	1214474123	Pusa 22	3212431122
Alankar	3234441142	Lee	3325422123	Punjab 1	3212451122
Ankur	1214541225	LSb1	1244544141	Pusa 40	1334524322
Birsa soybean 1	3221451122	MACS13	1314372341	PK416	1314121245
Bragg	1335544245	MACS57	3113551123	PS1042	1325651145
Co-1	1213531128	MACS58	3215444122	PS1024	1225414144
Co Soya-2	3323512123	MACS124	3112551122	PS 1225	1324414246
Co-3	1214633122	MACS450	1315524322	PS1029	1314154244
DS 228	3324414322	MAUS 1	1112554242	PK 308	3316164222
Pusa 97-12	1334114143	MAUS 2	1224672122	PK472	1224644145
Davis	1233374146	MAUS 61	3324421122	PK564	1335422236
Gujrat soybean-1	3214441122	MAUS 61-2	1224474142	PS1092	3124114122
Gujrat soybean-2	1311444121	MAUS 71	3324421144	PS1347	1214434145
Harit soy	1225274224	Monetta	3122331123	RBS2001-4	1325421122
Hardee	1215144122	MAUS 158	3324551121	RKS-24	1215444122
Indira soya 9	3313454125	MAUS47	3123451122	RAUS5	3323222144
Improved Pelican	3213254127	MAUS 81	3324421121	SL4	1314214243
JS 2	3221541122	MAUS 32	3214454125	Shilajeet	3324411124
JS20-34	1323344212	NRC 2	1124442125	SL96	3215341127
JS 71-05	3324421123	NRC 37	1215324122	PS1241	3125451143
JS72-44	1133341321	NRC 7	1334414227	SL 688	1314661245
JS 72-280	1434161221	NRC 12	1324424245	Shivalik	1314564127
JS 75-46	3223444227	NRC 86	1223324122	SL 295	1226254123
JS 76-205	1125154223	PRS 1	1114554236	SL 525	1336114241
JS79-81	3114541122	Pusa16	2315544222	Type 49	1115444122
JS 80-21	3225644227	Pusa20	3113474123	TAMS 38	3124454142
JS 90-41	1314444247	Pusa24	3313452346	TAMS 98-21	1312651345
JS 93-05	1334354227	Pusa37	2315451346	VLS1	1335124246
JS 95-60	1334454127	Pusa 98-14	1214444145	VLS2	1224444122
JS 97-52	1135222125	PK 262	1314144145	VL S21	1114422246
JS 335	3334411123	PS 471	1214425122	VLS47	1335142221
JS2029	1125224125	Palam soya	3221211242	VLS63	1214564128
Kalitur	1125154223	PK 327	3313444623	VLS 65	3425264123
KB 79	3222474142	Pratap Soya 2	1324411123	VLS59	1234444122

Each of the 10-digital DNA barcode represents the allele of primer-pair in the order (from left to right) of Satt538, Satt577, Satt267, Satt146, Satt352, Sct_199, Satt541, Satt181, Satt229 and Satt009

Three varieties, MACS124, MACS57 and MACS58, which have been developed from JS2 × Improved Pelican at the same breeding centre of All India Co-ordinated Project could be distinguished from each other through their respective DNA barcode. MACS124 and MACS57 were assigned barcodes 3112551122 and 3113551123, respectively, since these varieties showed the same alleles at eight loci; however, MACS124 and MACS57 showed 2nd (290 bp) and 3rd allele (300 bp) at Satt146 (4th position), respectively. These two varieties could also be distinguished from each other by amplicon of varying size i.e. 180 (2nd allele) and 190 bp (3rd allele) at Satt009 (10th position). MACS57 and MACS58 showed same digits (allele) at only four loci; while MACS58 (3215444122) and MACS124 (3112551122) inherited same alleles at five loci.

Gujrat soybean1 is a selection from Punjab1. The barcodes of Gujrat soybean-1 (3214441122) and Punjab1 (3212451122), showed that these two varieties can be distinguished from each other at loci Satt146 (4th digit) and Sct_199 (6th digit). Satt146 generated alleles of 310 and 290 bp, while Sct_199 generated allele of 210 and 220 bp in Gujrat soybean 1 and Punjab1, respectively. Pusa22 (3212431122) and Gujrat soybean1 (3214441122), both the varieties have Punjab1 as one of its parents, shared same alleles at eight loci and could be distinguished from each other by loci Satt146 and Sct_199, which generated alleles of different size in these two varieties. The barcode of Pusa22 (3212431122) was different from Punjab1 (3212451122) only at locus Sct_199, which generated an allele of 205 bp in Pusa22 and 220 bp in Punjab1. Similarly, Indira Soya 9 and MAUS32, both the varieties are selections from JS80-21. The barcode of JS80-21 (3225644227) was distinct from MAUS32 (3214454125) and Indira soya 9 (3313454125). MAUS32 (3214454125) and Indira soya 9 (3313454125) shared the same allele at eight loci. In these two varieties, different numerical numbers at the 2nd (Satt577) and 4th (Satt146) positions corresponding to different alleles at these loci in these two varieties.

The barcode of KB79 (3222474142) shared four alleles at Satt577, Satt541, Satt181and Satt009 with its parent Hardee (1215144122). PK416 is the parent of SL688. Barcodes of SL688 (1314661245) and PK416 (1314121245) showed same alleles at eight loci. PCR amplification using SSR markers Satt352 and Sct_199 distinguished these two varieties. Satt352 generated amplicons of 200 and 160bp; while Sct_199 generated amplicons of 225 and 200 bp in SL688 and PK416, respectively. VLS63 and VLS59, which have VLS2 as one of the parents, shared same alleles at six loci as indicated in their barcodes, while variety VLS2 (1224444122) can be distinguished from VLS59 (1234444122) only by Satt267 (3rd position), which generated an allele of 230 bp in VLS2 and 240 bp in VLS59. Further, three varieties NRC2, NRC12, VLS1 are mutant of Bragg. NRC2 (1124442125) differed from NRC12 (1324424245) and VLS1 (1335124246) at five and nine loci, respectively, while NRC12 differed from VLS1 at four loci.

JS72-44 and MACS13 has EC7034 as one common parent, while varieties Pusa40, Pusa16, Pusa20 and JS71-05, have Lee as one common parent. Similarly, a set of five varieties, KB79, PK308, PK262, PK471 and PK472, have Hardee as one common parent and another set of four varieties, PS1225, PS1092, PK1241 and JS97-52, has PK327 as one parent in its pedigree. JS335 (3334411123) is the common parent in four varieties, RAUS5, MAUS81, CO-3 and Co Soya-2; its barcode exhibited the maximum number

of common alleles with MAUS81 (3324421121) at seven loci, followed by Co soya-2 (3323512123) at six loci. Bragg (1335544245) is the common parent in 14 varieties; however, its barcode shared maximum digits with the barcodes of NRC12 (1324424245) at six loci and VLS1 (1335124246) at seven loci. The parent-progeny relationship reflected in some of the varieties through the 10-digit code in our study is in agreement with an earlier study in soybean Gao e*t al*. (2009) and other crops. Pilinsky *et al*. (2011) characterised 24 rapeseed varieties and hybrids using 12 SSR markers and proved the parental-progeny relationship of the new hybrids. Galbacs *et al*. (2009) converted SSR amplified data into barcodes in grape varieties and studied the parent-progeny relationship of genetic stock of unknown pedigree.

According to Brown *et al*. (1996), only 5-6 SSR markers with average PIC value are sufficient to distinguish 100 varieties. Gao *et al*. (2009) differentiated 83 soybean varieties cultivated in Heilongjiang province of China using nine SSR markers and generated a 9-digit molecular identity code. In our study, we used 10 SSR primer pairs to distinguish 102 soybean varieties. In this study, six specific alleles were detected in five varieties, while our study showed four specific alleles in four varieties. In both the studies, the banding pattern at the selected 9 or 10 loci were unique to each variety. The 10-digit code could distinguish varieties like MACS57, MACS58 and MACS124, developed from the common parents. The molecular identity code constructed for 83 soybean varieties in the study of Gao *et al*. (2009) could distinguish varieties with similar parentage and developed from the same institute. Vicario *et al*. (2001) differentiated 100 commercially cultivated soybean varieties of Argentina using 30 SSR markers, which was higher than the 10 primer pairs used in our study and nine in the study carried out by Gao *et al*. (2009). Efforts are underway to develop similar barcodes using SSR markers to distinguish varieties of other crops. Zhao *et al*. (2012) utilised 26 SSR markers to establish DNA barcodes for12 cotton cultivars.

Grow-out-tests, generally used to test the genetic purity of soybean seed samples of commercial varieties, are costly, tedious, time-consuming and affected by environment. DNA fingerprints, to generate the 10-digit barcode developed in this study, would be simpler, faster and more accurate for testing genetic purity of seeds lots, drawn from a particular variety. As this technique uses simple PCR based amplification and amplicons generated by SSR markers are highly reproducible, and do not vary across the laboratories. Various studies have proven the use of SSR markers in genetic purity testing in other crops. Use of SSR markers in genetic purity testing has been successfully demonstrated in broccoli (Yu *et al.,* 2013) and rice hybrids (Bora *et al.,* 2016).

To conclude amplicons profile generated by 10 SSR markers were converted into barcodes for 102 varieties. These unique identification codes explained the pedigree of closely related soybean varieties and in combination with DUS (distinctness, uniformity and stability) testing; they can prove to be very effective in accurately analysing the new candidate varieties of soybean and protecting the plant breeders' rights.

Acknowledgements

The authors acknowledge the Indian Council of Agricultural Research for facilitating this research work under the in-house project.

References

Ahemed, M.M., Guo, H., Huang, C., Zhang, X. and Lin, Z. (2013). Selection of core SSR markers for finger-printing upland cotton cultivars and hybrids. *Australian Journal of Crop Science*, **7**, 1912-1920.

Bora, A., Choudhury, P.R., Pande, V. and Mandal, A.B. (2016). Assessment of genetic purity in rice (*Oryza sativa* L.) hybrids using microsatellite markers. *3 Biotech*, **6**, 1-7.

Brown, S.M., Hopkins, M.S., Mitchell, S.E., Senior, M.L., Wang, T.Y., Duncan, R.R., Gonzalez-Candelas, F. and Kresovich, S. (1996). Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum. *Theoretical and Applied Genetics*, **93**, 190-198.

Community Reference Laboratory for GM Food and Feed. (2007). Soybean seeds sampling and DNA extraction. In *Report on the Validation of a DNA Extraction Method for Soybean Seeds*, European Commission Directorate-General, Joint Research Centre, Institute of Health and Consumer Protection, Biotechnology and GMO Unit, pp. 1-14.

Galbacs, Z., Molnar, S., Halasz, G., Kozma, P., Hoffmann, S., Kovacs, L., Veres, A., Galli, Z., Szőke, A., Heszky, L. and Kiss, E. (2009). Identification of grapevine cultivars using microsatellite-based DNA barcodes. *Vitis*, **48**, 17-24.

Gao, Y.-L., Zhu, R.-S., Liu, C.-Y., Li, W.-F., Jiang, H.-W., Li, C.-D., Yao, B.-C., Hu, G.-H. and Chen, Q.-S. (2009). Constructing molecular identity for soybean varieties from Heilongjiang Province, China. *ACTA Agronomica Sinica*, **35**, 211-218.

Joshi, N., Sharma, S., Subramanian. and Rao, K.S. (2013). Genetic DNA fingerprinting of chick pea germplasm using morphological and molecular markers. *Asian Journal of Experimental Biology Science*, **4**, 398-405.

Khan, A.Z., Shah, P., Khalil, S.K. and Taj, F.H. (2003). Influence of planting date and plant density on morpho-logical traits of determinate and indeterminate soybean cultivars under temperate environment. *Journal of Agronomy*, **2**, 146-152.

Kumar, V., Rani, A., Jha, P., Rawal, R. and Hussain, S.M. (2015). Molecular identification of dominant cultivars of soybean using simple sequence repeats markers. *Proceedings of National Academy of Sciences, India, Section B Biological Sciences,* DOI 10. 1007/s40011-015-0641-7.

Novakova, A., Simackova, K., Barta, J. and Curn, V. (2010). Utilization of DNA markers based on microsatellite polymorphism for identification of potato varieties cultivated in Czech Republic. *Journal of Central European Agriculture*, **11**, 415-422.

Pilinsky, K., Falusi, J., Szoke, A., Kiss, E. and Hesszky, L. (2011). Characterization of oilseed genotypes using microsatellite based DNA barcode. In the *Proceedings of 13th International Rapeseed Congress*, Organized by International Consultative Group of Research on Rapeseed 5-9 June, 2011, Prague, Czech Republic.

Sudharani, M., Rao, P.S. and Subba Rao, L.V. (2014). Identification of SSR markers for testing of hybridity and seed genetic purity in maize. *International Journal of Scientific Research*, **3**, 92-95.

Tantasawat, P., Trongchuen, J., Prajongjai, T., Jenweerawat, S. and Chaowiset, W. (2011). SSR analysis of soybean (*Glycine max* (L.) Merr.) genetic relationship and variety identification in Thailand. *Australian Journal of Crop Sciences*, **5**, 283-290.

UPOV/INF/17/1 (2010). Guidelines for DNA-profiling: Molecular marker selection and database construction. In *Proceedings of the 44th Session of International Union for the Protection of New Varieties of Plants*, 21, October, pp. 1-13, Geneva.

Vicario, A., Giancola, S.M., Echaide, M. and Hopp, H.E. (2001). Application of micro-satellite markers for the assessment of distinctness, uniformity and stability (DUS) testing of commercial soybean varieties. In *Proceedings of* 7th Session of *International Union for the Protection of New Varieties of Plants*, 12 Nov, 2001, pp. 1-10, Hanover, Germany.

Yu, H.F., Wang, J.S., Zhao, Z.Q., Sheng, X.G. and Gu, H.H. (2013). DNA fingerprinting and genetic purity testing of a new broccoli hybrid using SSR markers. *Seed Science and Technology*, **41**, 464-468. <http://doi.org/10.15258/sst.2013.41.3.13>

Zhang, Y., Dai, S., Hong, Y. and Song, X. (2014). Application of genomic SSR locus polymorphisms on the identification and classification of chrysanthemum cultivars in China. *PLOS ONE*, **9**, e104856.

Zhao, L., Cai, C.P., Mei, H. X. and Guo, W.Z. (2012). Screening of microsatellite loci for identifying genome barcoding of cotton cultivars. *ACTA Agronomica Sinica*, **38**, 1810-1817.

Zhu, Y., Hu, J., Han, R., Wang, Y. and Zhu, S. (2011). Fingerprinting and identification of closely related wheat (*Triticum aestivum* L.) cultivars using ISSR and fluorescence-labeled TP-M13-SSR markers. *Australian Journal of Crop Science*, **5**, 846-850.

Zhu, Y.F., Qin, G.C., Hu, J., Wang, Y., Wang, J.C. and Zhu, S.J. (2012). Fingerprinting and variety identification of rice (*Oryza sativa* L.) based on simple sequence repeat markers. *Plant Omics*, **5**, 421-426.