RESEARCH ARTICLE

SSR analysis of 38 genotypes of soybean (Glycine Max (L.) Merr.) genetic diversity in India

Anchal Bisen · Dhirendra Khare · Priya Nair · Niraj Tripathi

Received: 1 September 2014 /Revised: 12 October 2014 /Accepted: 2 November 2014 /Published online: 6 December 2014 \odot Prof. H.S. Srivastava Foundation for Science and Society 2014

Abstract Sixteen polymorphic Simple sequence repeat (SSR) markers were used to determine the genetic diversity and varietal identification among 38 soybean (Glycine max (L.) Merr.) genotypes which are at present under seed multiplication chain in India. A total of 51 alleles with an average of 2.22 alleles per locus were detected. The polymorphic information content (PIC) among genotypes varied from 0.049 (Sat 243 and Satt 337) to 0.526 (Satt 431) with an average of 0.199. The pair wise genetic similarity between soybean varieties varied from 0.56 to 0.97 with an average of 0.761. These 16 SSR markers successfully distinguished 12 of the 38 soybean genotypes. These results suggest that used SSR markers are efficient for measuring genetic diversity and relatedness as well as identifying varieties of soybeans. Diverse genetic materials may be used for genetic improvements of soybean genotypes.

Keywords Genetic diversity · Molecular markers · Soybean · Varietal identification

Abbreviations

A. Bisen \cdot D. Khare $(\boxtimes) \cdot$ P. Nair \cdot N. Tripathi Seed Technology Centre, Jawaharlal Nehru Agriculture University, Jabalpur, M.P., India 482004 e-mail: dhirendrakhare@gmail.com

A. Bisen : D. Khare : P. Nair : N. Tripathi Biotechnology Centre, Jawaharlal Nehru Agriculture University, Jabalpur, M.P., India 482004

Introduction

The genetic diversity is of great significance for planning an efficient breeding programme for the improvement in any crop (Chandra et al. [2013\)](#page-6-0). Soybean (Glycine max (L.) Merr.) is one of the world's most important economic legume crops. Thousands of breeding lines and number of varieties are developed yearly through soybean hybridization programmes all over the world. At present, the number of extant variety in India is >100. Intensive breeding is engaged with increased genetic uniformity in the frame of species. Generations of new and improved cultivars can be enhanced by new sources of genetic variation; therefore criteria for parental stock selection need to be considered not only by agronomic value, but also for genetic dissimilarity. Therefore, evaluation of genetic variation is a very important task not only for population genetics but also for plant breeders. To protect a variety under Protection of Plant Variety and Farmers Right Act 2001, the candidate variety should have distinctiveness from other varieties, uniformity in expression of distinguishing traits among the plants in the population and stability in expression of distinguishing traits over the location and year.

Evaluation of genetic divergence and relatedness among breeding materials has significant implications for the improvement of crop plants. Knowledge on genetic diversity in soybean could help breeders and geneticists to understand the structure of germplasm, predict which combinations would produce the best offspring and facilitate to widen the genetic basis of breeding material for selection. In India, soybean is known for a low level of genetic diversity base in morphological expression of distinguishing traits due to narrow genetic base. It makes examination of distinctiveness difficult. Among various markers available for genetic analysis in plants, molecular markers are more efficient, precise and reliable in discriminating closely related species and cultivars. Simple sequence repeats (microsatellites) are sequences of short tandem repeats distributed over the genomes. The hyper-variable number of repeat sequences makes them an excellent tool for genotype differentiation, pedigree analysis, evaluation of genetic distances among organisms and identification of particular genotype etc. In particular, SSRs have been used successfully in identification of genetic diversity and relationships among soybean genotypes in different populations (Guan et al. [2010](#page-6-0); Wang et al. [2010\)](#page-6-0).

The present investigation was carried out with an objective to study the diversity level among the genotypes. Identification of marker for particular genotype was another thrust area of the present study. Genetic distances will further help in identifying genetically diverse genotypes, which then can be utilized in creating valuable selectable variation.

Materials and methods

The plant material comprises of soybean varieties in active seed multiplication chain along with all the varieties developed and released by JNKVV, Jabalpur. The seeds were obtained from the institutes developed the variety (Table 1).

DNA isolation

Genomic DNA was isolated from bulk young leaves of 10 plants from each genotype following the Cetyl Trimethyl Ammonium Bromide (CTAB) method of Saghai-Maroof et al. [\(1984\)](#page-6-0). DNA was quantified by comparing the intensity of ethidium bromide-stained DNA bands on 0.8 % agarose gels. Final concentration was adjusted to 50 ng μ ⁻¹ for further uses in PCR analysis.

Simple sequence repeats (SSR) analysis

A total of 50 SSRs primer pairs, distributed across the integrated linkage map of soybean (Cregan et al. [1999\)](#page-6-0) were used and synthesized by Integrated DNA Technologies. Most of the SSR markers that are included had an (ATT)n motif due to their abundance and polymorphic nature in soybeans and their easily interpretable allele patterns (Narvel et al. [2000\)](#page-6-0). DNA was amplified by PCR using our previously standardized method (Sahu et al. [2012](#page-6-0)) in a total volume of 10 μl containing 25 ng template DNA, $1 \times$ buffer (75 mM Tris.HCl (pH 9.0), 50 mM KCl, 20 mM (NH4)2SO4), 2 mM MgCl2,

200 μM of each dNTP, 5 pmol SSR primers, and 1 unit DNA polymerase (Fermentas). PCR reactions were carried out in an ESCO thermocycler. Cycling parameters were initial denaturation step at 94 °C for 5 min, followed by 94 °C, 30 s, 52–58 °C, 30 s and 72 °C, 30 s. This cycle was repeated 35 times, followed by 5 min final extension at 72 °C. The amplified products were separated on 3.5 % agarose gels and detected by ethidium bromide staining.

Allele sizes were estimated in comparison with 100 bp DNA ladder (Fermentas).

Data scoring

The PCR products from SSR were analyzed by scoring qualitatively for presence or absence. A genetic similarity between the cultivars was measured by the similarity coefficient. In case of SSRs, Polymorphism Information Content (PIC) was calculated according to Anderson et al. [\(1993](#page-6-0)) using the following equation: $\text{PIC}_j = 1 - \sum_{i=1}^{n} P_i^2$ Where, $i =$ the ith allele ronowing equation: $P1C_j = 1 - \sum_{i=1} P_i$ where, $1 = \text{the } 1$ allele
of the jth marker, n = the number of alleles at the jth marker and $p =$ allele frequency.

Results and discussion

The assessment of genetic diversity is not only important for crop improvement efforts but also for the efficient management and protection of available genetic variability. Molecular profiling has been the preferred choice of breeding as these are more reliable & authentic and less influenced by environmental fluctuations (Vinu et al. [2013\)](#page-6-0). The molecular marker in the other hand can be utilized for identification of marker for specific trait. Such diversity studies are helpful in categorizing the population into diverse group and that will help in development of gene pool.

A total of 50 SSR primer pairs, distributed on 10 of 20 linkage groups of soybean (Cregan et al. [1999\)](#page-6-0), were used to amplify specific SSR loci from bulked DNA of each soybean genotype. Among these SSR primers, only 23 markers lacated on 9 linkage groups were amplified scorable loci. Sixteen primer pairs produced clear single-locus polymorphic bands for the analysis (Table 2) and remaining seven were Monomorphic. The high percentage of polymorphic SSR loci (78.4 %) detected in this study was consistent with previous studies (Maughan et al. [1995](#page-6-0); Rongwen et al. [1995](#page-6-0); Diwan and Cregan [1997;](#page-6-0) Narvel et al. [2000;](#page-6-0) Kumar et al. [2009;](#page-6-0) Singh et al. [2010\)](#page-6-0). A total of 51 alleles from 23 SSR markers were detected across all 38 genotypes. The number of alleles per primer pair (locus) ranged from 1 for all monomorphic markers (Satt183, Satt229, Satt269, Satt294, Satt326, Satt399 and Sct_187) to 5 for Satt 373 and Satt 431 with an average of 2.21 (Table [3](#page-3-0)).

Unique alleles specific for PS 1241, JS90-41, JS93-05, MACS450, JS72-280, JS75-46, Harasoya, NRC2, JS97-52, JS76-205, JS20-34 and Bragg were observed in this study and

Table 2 SSR markers with their sequences selected for the study [\(http://www.soybase.org\)](http://www.soybase.org/)

Sr.	Primer	Forward sequence (5'-3')	Reverse sequence (5'-3')	Linkage group
1	Sat 183	GCGTCCAGCCTGACCATTTTA	GCGTTCCAATGTCTGATTATTT	D1h
2	Sat 243	GCGATGTCGAATGATTATTAATCAAAATC	GCGGCAACCGCTTAAAAATAATTTAAGAT	K
3	Sat_366	GCGGCACAAGAACAGAGGAAACTATT	GCGGACATGGTACATCTATATTACGAGTATT	J
4	Satt177	CGTTTCATTCCCATGCCA ATA	CCCGCATCTTTTTCAACCAC	A2
5	Satt229	TGGCAGCACACCTGCTAAGGGAATAAA	GCGAGGTGGTCTAAAATTATTACCTAT	\mathbf{L}
6	Satt244	GCGCCCCATATGTTTAAATTATATGGAG	GCGATGGGGATATTTTCTTTATTATCAG	J
7	Satt245	AACGGGAGTAGGACATTTTATT	GCGCCTCCTGAATTTCAAAGAATGAAGA	M
8	Satt264	CCTTTTGACAATTATGGCATATA	GCATAGAAGGGCATCATTCAGAT	K
9	Satt269	GCGTGCCAGGTAGAAAAATATTAG	GCGGTTTTTCACTTTTCAAAATTC	K
10	Satt285	GCGACATATTGCATTAAAAACATACTT	GCGGACTAATTCTATTTTACACCAACAAC	J
11	Satt288	GCGGGGTGATTTAGTGTTTGACACCT	GCGCTTATAATTAAGAGCAAAAGAAG	G
12	Satt294	GCGGGTCAAATGCAAATTATTTTT	GCGCTCAGTGTGAAAGTTGTTTCTAT	C ₁
13	Satt308	GCGTTAAGGTTGGCAGGGTGGAAGTG	GCGCAGCTTTATACAAAAATCAACAA	M
14	Satt326	AGATTCTCCTTTGCTTCTTAGT	GTTAGTTCACCTTCCAGTATTTGA	K
15	Satt ₃₃₇	GCGTAAATCTGATATATGTTACCACTGA	GCGTAATACGCAAAACATAATTAGCCTA	K
16	Satt373	TCCGCGAGATAAATTCGTAAAAT	GGCCAGATACCCAAGTTGTACTTGT	\mathbf{L}
17	Satt399	AAGCCAACCTTATAATTCTTTCAT	ATATGGGCTTACTTACCCATCATAGA	C ₁
18	Satt406	GCGTGAGCATTTTTGTTT	TGACGGGTTTAATAGCAT	J
19	Satt414	GCGTATTCCTAGTCACATGCTATTTCA	GCGTCATAATAATGCCTAGAACATAAA	J
20	Satt431	GCGTGGCACCCTTGATAAATAA	GCGCACGAAAGTTTTTCTGTAACA	J
21	Satt440	TGAGAACGTTTGAAAAGAGAT	GAAGAGATTAAGCATAAAGAATACTT	T
22	Satt562	GCGGATTCACTAGGATGTTTAT	GCGGCGGCAGCTTAAATGGATTGA	I
23	Sct_187	CATGCTCCCATTCTCT	AACATTGGCTTTTTACTTAG	G

Table 3 Major allele frequency, genetic diversity, heterozygosity and PIC value of the amplified SSR markers in soybean

may be useful for DNA fingerprinting. Presence of unique band helped in the identification of specific genotype. Such markers are highly reliable in the establishment of genetic relatedness among the genotypes. Similar results were reported by Jain et al. ([1994\)](#page-6-0), Srivastava et al. [\(2001](#page-6-0)), and Vinu et al. [\(2013](#page-6-0)) in different crop species. PIC value is a measure of the allelic differentiation. The highest PIC value was observed for the marker Satt431 (0.5263) and the lowest (0.0499) for the markers Sat 243 and Satt337 with the average PIC value of 0.1998 (Table 3). Eight SSR markers (Satt177, Satt245, Satt264, Satt288, Satt373, Satt406, Satt431 and Satt440) had PIC value greater than 0.3 and hence were the most informative for distinguishing among the soybean genotypes. These markers occurred on 7 separate LGs, indicating that molecular polymorphism was spread across different regions of the genome. The SSR diversity detected among soybean genotypes in this study was moderate compared to that from most previous reports. Diwan and Cregan ([1997](#page-6-0)) detected an average of 10.10 alleles per locus and an average marker diversity of 0.80 when 20 SSR markers were used to distinguish the 35 North American soybean genotypes. Narvel et al. [\(2000\)](#page-6-0) calculated an average of 10.20 alleles per locus among 39 soybean elite genotypes and 40 PIs from seven different countries using 74 SSR markers.

The study revealed that the allele frequency was highest (0.9737) in Sat_243 and Satt337 and lowest in Satt288 (0.4474) with the mean of 0.8244. Only one marker Satt288 had allele frequency less than 0.50(0.4474). Genetic diversity showed the probability that two randomly chosen alleles are different from the population varied from 0.0512 (Sat 243, Satt337) to 0.6053 (Satt288) with an average of 0.2339. The heterozygosity among the 23 amplified SSR markers was low. Only four markers i.e., Satt373, Satt406, Satt431 and Satt440 showed heterozygosity that ranged between 0.0263 (Satt440) to 0.7368 (Satt406) with an average of 0.0469 (Table 3). Unique alleles were obtained in eight markers viz., Sat_243, Sat_366, Satt285, Satt337, Satt373, Satt414, Satt440 and Satt562. These markers can be used to discriminate particular genotype. The highest range (310–400 bp) of allele size was found in Satt 244. Whereas, lowest (150–180 bp) in Satt288 (Table [4\)](#page-4-0).

Genetic relationships among soybean genotypes

Similarity matrices of all the 38 varieties were generated using Power Maker version 3.25 (Liu and Muse [2005](#page-6-0)). The allelic diversity data was used to produce a dendrogram (cluster tree analysis, Power Maker) to explain the genetic relationships among the varieties (Fig. [1\)](#page-5-0). All genotypes were grouped into

two clusters. Cluster I was the largest and the most diverse cluster consisting of 36 varieties. This cluster was further divided into 2 sub-groups; sub-group 'A' contains 28 varieties and sub-group 'B' 8 varieties. Cluster II includes two varieties. The first tertiary group has 28 varieties of different origin viz. Bragg, Harasoya, Pratap Soya, NRC2, NRC7, NRC37 and PUSA9814. Variety Pratap Soya 2 and PS1042 are very closely related (0.975). The second tertiary groups constituted of varieties MAUS71, MAUS81, MAUS158, MAUS61, NRC7, NRC2, NRC37, PUSA9814, PUSA9712, SL525, TAMS38, and VLS47. This subgroup A consisted of seven varieties viz. Harasoya, MACS 450, PS 1042, Pusa 9814, JS 2, JS 72–280, JS 76–205 having common parent 'Bragg'. From these seven varieties, JS76-205 and JS72-280, Pusa 9814 and PS 1042 showed highest (87.80 %) whereas, Harasoya and JS 72–280 showed least (58.53 %) similarity. Three varieties viz. PS1042, PS1092 and SL525 have one common parent PK 416 in their pedigree, out of which PS1042 and PS 1092 showed 92.68 % similarity. Further in this sub-group A, TAMS 38 and PS 1347 had 90.24 % similarity with common parent PK 472. Also TAMS 38 and Pratap Soya 2 as well as MAUS 71 and JS335 showed 80.48 and 73.17 % similarity along with common pedigree parent Monetta and JS 71–05, respectively. The second subgroup 'B' consists of JS90-41, MAUS61, JS97-52, MACS450, JS93-05, JS95-60, Pratap Soya 1 and JS72-44. Here, JS95-60 and JS93-05 are derived from same pedigree (Secondary selection from PS 73–22) showed 90.24 % similarity. Only two varieties PS 1225 and PS 1241 representing minor group were found to be extremely diverse from the other varieties under study. The cluster analysis revealed that the varieties viz., PS1225 and PS1241 can be utilized in breeding programme due to its diversity and uniqueness from other varieties. The pairwise genetic similarity between soybean varieties varied from 0.56 (JS 2 v/s NRC 2 and JS72-280 v/s PS1241) to 0.97 (Pratap Soya 2Vs PS 1042) with an average of 0.761. This was in dichotomy with the result of Priolli et al. ([2002\)](#page-6-0), in which 12 SSR loci were used to distinguish morphologically similar groups, depicting a mean similarity coefficient of 0.46; Singh et al. ([2010](#page-6-0)), where the similarity coefficient based on shared SSR and AFLP bands revealed that the genetic similarity ranged from 0.220 to 0.765 with an average of 0.446. Further, the result of the present investigation was in disagreement with the finding of Tantasawat et al. ([2011](#page-6-0)) who reported pairwise genetic similarity to vary from 0.73 to 1.0 with an average of 0.88. This

Fig. 1 Dendrogram showing diversity among 38 varieties of soybean based on SSR markers

low level of genetic diversity may be ascribed to the emphasis on direct introductions, selection from introduced germplasm and single cross hybrids (some of which shared common parents) in the soybean breeding programs. Therefore, inclusion of more diverse germplasm in the soybean breeding programs may provide the genetic variability necessary to permit continued progress and broad adaptation.

The unique or rare alleles caused by natural mutation and selection (Mousadik and Petit [1996](#page-6-0)) can be used not only in the specific categorization of germplasm collections, but also in their subsequent utilizations in breeding and plant development as unique markers. In the present study, ten unique alleles were amplified by eight SSR (Sat_243, Sat_366, Satt285, Satt337, Satt373, Satt414, Satt440 and Satt562) markers specific for 12 varieties i.e., Bragg, Harasoya, JS20- 34, JS72-280, JS75-46, JS76-205, JS90-41, JS93-05, JS97- 52, MACS450, NRC2 and PS1241 (Table 5) were observed. These unique alleles can be utilized for the identification of specific distinguishing genes/alleles for further studies. Similarly, Tantasawat et al. [\(2011\)](#page-6-0) and Sahu et al. [\(2012\)](#page-6-0) obtained seven unique alleles in their study on soybean genotypes using SSR markers. Similarly, SSR markers were used for varietal identification with trait association in rice (Reddy et al. [2013\)](#page-6-0). The SSR primers, particularly when used together with multiplex technology, could be useful for accurate and cost-effective genotyping for soybean variety identification.

The present investigation resulted with the markers differentiating soybean varieties can be exploited in molecular breeding by combining with other useful genes to improve soybean varieties. Information on genetic distances based on microsatellite markers shall be preferred in creating selectable genetic variation using genotypes which are genetically apart (Vieira et al. [2007](#page-6-0); Vinu et al. [2013\)](#page-6-0). The diversity analysis can further be utilized for the development of diverse gene pool. The hybridization among the diverse gene pool will result into more heterotic combinations.

Table 5 Allele size of SSR markers in soybean varieties with unique alleles

Marker		Unique Allele Allele size (bp)	Genotype I	Genotype II
Sat 243	1	390	PS1241	NRC ₂
Sat 366	2	230	JS90-41	PS1241
		200	JS93-05	JS97-52
Satt285	1	235	JS90-41	JS76-205
Satt337	1	245	MACS450	
Satt373	2	210/225	MACS450	
		190/225	JS72-280	JS76-205
Sat414	1	240	MACS450	JS20-34
Satt440	1	185/200	JS75-46	
Satt ₅₆₂	1	210	Harasoya	Bragg

References

- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. Genome 36:181–186
- Chandra V, Pant U, Bhajan R, Singh AK (2013) Studies on genetic diversity among Alternaria blight Tolerant Indian mustard genotypes using SSR markers. The Bioscan 8(4):1431–1435
- Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark KG, Kahler AL, Kaya N, vanToai TT, Lohnes DG, Chung J, Specht JE (1999) An integrated genetic linkage map of the soybean genome. Crop Sci 39: 1464–1490
- Diwan N, Cregan PB (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor Appl Genet 95:723–733
- Guan R, Chang R, Li Y, Wang L, Liu Z, Qiu L (2010) Genetic diversity comparison between Chinese and Japanese soybeans (Glycine max (L.) Merr.) revealed by nuclear SSRs. Genet Resour Crop Evol 57: 229–242
- Jain A, Bhatia S, Banga SS, Prakash S, Lakshmikumaran M (1994) Potential use of random amplified polymorphic DNA (RAPD) technique to study the genetic diversity in Indian mustard (Brassica juncea) and its relationship to heterosis. Theor Appl Genet 88:116–122
- Kumar J, Verma V, Goyal A, Shahi AK, Sparoo R, Sangwan RS, Qazi GN (2009) Genetic diversity analysis in Cymbopogon species using DNA markers. Plant Omics J 2:20–29
- Liu K, Muse M (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129
- Maughan PJ, Saghai-Maroof MA, Buss GR (1995) Microsatellite and amplified sequence length polymorphism in cultivated and wild soybean. Genome 38:715–723
- Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [Arganiaspinosa (L.) Skeels] endemic to Morocco. Theor Appl Genet 92:832–835
- Narvel JM, Fehr WR, Chu WC, Grant D, Shoemaker RC (2000) Simple sequence repeat diversity among soybean plant introductions and elite genotypes. Crop Sci 40:1452–1458
- Priolli RHG Jr, Mendes CT, Arantes NE, Contel EPB (2002) Characterization of Brazilian soybean cultivars using microsatellite markers. Genet Mol Biol 25:185–193
- Reddy ESS, Verma SK, Xalxo SM, Saxena RR, Verulkar SB (2013) Identification of molecular marker(s) for root Length in rice (oryza sativa l.). The Bioscan 8(4):1511–¹⁵¹⁴
- Rongwen J, Akkaya MS, Bhahwat AA, Lavi U, Cregan PB (1995) The use of microsatellite DNA markers for soybean genotype identification. Theor Appl Genet 90:43–48
- Saghai-Maroof K, Soliman M, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci U S A 81:8014–8018
- Sahu P, Khare D, Tripathi N, Shrivastava AN, Saini N (2012) Molecular screening for disease resistance in soybean. J Food Leg 25(3):200– 205
- Singh RK, Mishra SK, Singh SP, Mishra N, Sharma ML (2010) Evaluation of microsatellite markers for genetic diversity analysis among sugarcane species and commercial hybrids. Aus J Crop Sci 4:116–125
- Srivastava A, Gupta V, Pental D, Pradhan AK (2001) AFLP based genetic diversity assessment amongst agronomically important natural and some newly synthesized lines of Brassica juncea. Theor Appl Genet 102:193–199
- Tantasawat P, Trongchuen J, Prajongjai T, Jenweerawat S, Chaowiset W (2011) SSR analysis of soybean (Glycine max (L.) Merr.) genetic relationship and variety identification in Thailand. Aus J Crop Sci 5(3):283–290
- Vieira E, Carvalho F, Bertan I, Kopp M, Zimmer P, Benin G, da Silva J, Hartwig I, Malone G, de Oliviera A (2007) Association between genetic distances in wheat (Triticum aestivum L.) as estimated by AFLP and morphological markers. Genet Mol Biol 30:392–399
- Vinu V, Singh N, Vasudev S, Yadava DK, Kumar S, Naresh S, Bhat SR, Prabhu KV (2013) Assessment of genetic diversity in Brassica juncea (Brassicaceae) genotypes using phenotypic differences and SSR markers. Rev Biol Trop 61(4):1919–1934
- Wang M, Li RZ, Yang WM, Du WJ (2010) Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. Afr J Biotechnol 9:4857–4866