

Understanding genetic relationship and population structure of Indian soybean varieties using microsatellite markers

Devendra Kumar Chauhan¹  · Javaid Akhter Bhat¹ · Ajay Kumar Thakur² · Zakir Hussain³ · C. Tara Satyavathi⁴

Received: 12 May 2016/Revised: 9 November 2016/Accepted: 12 January 2017/Published online: 3 February 2017
© The National Academy of Sciences, India 2017

Abstract Thirty-nine simple sequence repeat markers have been used to estimate the genetic diversity and population structure of 78 soybean varieties released in India. A total of 140 alleles were detected and an average of 3.59 alleles per locus was recorded. Polymorphism information content values for all the polymorphic markers across 78 soybean varieties ranged from 0.073 (Satt415) to 0.931 (Satt084), with an average value of 0.59. Mean similarity coefficient was recorded as 0.41 for all the genotypes. Marker index values ranged from 0.051 (Satt498) to 5.556 (SOYSHPI76) with an average value of 2.54. The resolving power value ranged from 0.102 (satt498) to 2.086 (GMABABE) with an average value of 1.64. Both the phylogenetic and population structure analysis classified soybean varieties into five main clusters and five sub-populations, respectively. Principal coordinate analysis results are in agreement with both of these methods, which grouped the varieties on the basis of their origin, pedigree and releasing centre.

Keywords Genetic diversity · Soybean · Simple sequence repeat markers · Polymorphism

Introduction

Soybean (*Glycine max* L. Merrill; $2n = 40$), a “miracle legume” native from China, is currently one of the most important crop worldwide. In India, this is the major contributor of edible oil after groundnut, with a cultivated area of 10 million ha and 10.44 million ton production [1]. However, the average productivity of this crop is one ton/ha, which is very low to meet the ever increasing demand of edible oil in India. Development of improved and new varieties is the main aspect for increasing the overall production and productivity of this very important oilseed crop. In India, 78 varieties have been developed from various breeding centres and this figure is increasing every year, but no significant improvement in yield or resistance/tolerance to various biotic and abiotic stresses has been achieved so far. After the implementation of “Protection of Plant Varieties and Farmers Rights Act-2001”, varietal identification got much more attention. When referring to the necessary requirement for the protection of new cultivars, it states that the cultivar should be reliably “New, Distinct, Uniform and Stable”. Along with the development of new varieties, there has been a growing interest in genetic characterization for protection of varieties provided by PPV and FR Act. Also for the breeding and genetic improvement of crop species, a thorough knowledge of the genetic diversity is necessary to decide the breeding strategies involving the selection of parents to maximize genetic improvement. More accurate and complete descriptions of existing soybean varieties and patterns of genetic diversity could facilitate introgression of diverse germplasm into current commercial soybean genetic base.

Plant breeders have traditionally used morphological and biochemical markers for the genetic diversity analysis and the registration of varieties under PPV and FR Act.

✉ Devendra Kumar Chauhan
chauhanskuastj@gmail.com

¹ Present Address: Division of Plant Breeding and Genetics, Sher-e-Kashmir University of Agricultural Sciences and Technology of Jammu, Chatha, Jammu, J&K 180009, India

² ICAR-Directorate of Rapeseed-Mustard Research, Bharatpur, Rajasthan 321303, India

³ Division of Vegetable Sciences, ICAR-Indian Agricultural Research Institute, Pusa, New Delhi 110012, India

⁴ Division of Genetics, ICAR-Indian Agricultural Research Institute, Pusa, New Delhi 110012, India

These are limited in numbers, highly influenced by the environment and are growth stage-specific. These traits create difficulty particularly in closely related cultivars. In plants with a narrow genetic base in their gene pool such as soybean, they may not be sufficient for the diversity analysis and identification of genotypes. So a more reliable and efficient method is required for characterisation and identification of genotypes. In such cases, molecular markers can provide additional information about the genetic characterisation, degree of diversity and genetic relationship of the existing genotypes [2]. Plant protection office of USDA Agriculture Marketing Services now accepts microsatellite allelic profiles as supporting evidence for the uniqueness of a cultivar [3].

Among the class of molecular markers, microsatellite markers/simple sequence repeats (SSRs) are the sequences of a few repeated and adjacent base pairs, which are very useful for variety of applications in plant genetics and breeding. There are a number of advantages of using SSRs over various other molecular markers including their co-dominant and multi-allelic nature [4], less expensive PCR-based assay, relative abundance, uniform distribution over the eukaryotic genome [5] and high reproducibility, which are very suitable for accession discrimination and assessment of genetic variation. SSR markers have been useful for integrating the genetic, physical and sequence based physical maps in plant species and simultaneously have provided breeders and geneticist with an efficient tool to link phenotypic and genotypic variation [6]. Variation in the number of repeats can be detected by polymerase chain reaction (PCR) with the development of primers (20–30 base pairs), specifically designed for amplification and complementary to single sequences flanking the microsatellite. These markers have been used for studying the genetic relationship in soybean [7–17] and many other crops including rice [18], wheat [19], Indian mustard [20], faba bean [6] etc. However, no exhaustive study of genetic diversity, genetic relationship and population structure of Indian soybean cultivars had been carried out so far using SSR markers. The objective of the present study was to reveal the genetic relationship and population structure of Indian soybean varieties released from different soybean breeding centres using SSR markers.

Material and Methods

Plant Material and DNA Isolation

Seventy-eight soybean varieties developed by different breeding centres and released by central/state varieties release committees for different agro climatic regions were taken up for the present study. Detailed description of

pedigree and releasing centres has been shown in Table 1. Genomic-DNA was extracted from young and tender leaves of soybean varieties using the cetyl trimethyl ammonium bromide (CTAB) method [21] and purified by RNAase treatment. DNA quality and concentration were evaluated on 0.8% agarose gel electrophoresis stained with ethidium bromide in the presence of different concentrations of undigested λ -DNA and a final concentration of 50 ng μl^{-1} was used for PCR.

SSR Genotyping

Simple sequence repeats markers (39 primer pairs), previously developed [22], distributed across all the 20 linkage groups of soybean genome, with at least 2 markers from each linkage group were selected and custom synthesized (Table 2). Detailed description of the primers is available at soybean website USDA-ARS Soybean Genome Database (<http://129.186.26.94/SSR.html>). Primers were chosen because of their trinucleotide nature and presented polymorphism in the previous studies. The PCR reaction (25 μl) contained 1X reaction buffer (20 mM Tris-Cl, pH 8.4, 50 mM KCl), 10 nM dNTPs, 50 mM MgCl_2 , 5 pM primer, 1.0 Unit of *Taq* DNA polymerase and 50 ng genomic DNA. For standardization of annealing temperatures of SSR primers, gradient PCR was carried out in a gradient thermal cycler. Initial denaturation at 94 °C for 5 min was followed by 35 cycles at 94 °C for 2 min, 47 °C for 1 min and 72 °C for 1 min. The final extension was carried out at 72 °C for 10 min. The SSR amplified fragments were resolved in 3% metaphore agarose gel in a 1X TBE buffer. The gels were stained with ethidium bromide (0.5 $\mu\text{g}/\text{ml}$) and visualized under UV light.

Statistical Analysis

The prominent DNA bands that were amplified by a given primer, were scored as present (1) or absent (0) for all of the samples under investigation. To determine the utility of the SSR markers, number of amplicons per marker, Polymorphic Information Content (PIC), Major allelic frequency, Marker Index (MI), Resolving power (Rp) and Discrimination power (Dp) were calculated for each pair of primers. The Polymorphism Information Content (PIC) value of individual primer pair was calculated based on the following formula [23].

$$PIC = 1 - \sum_{i=1}^n P_{ij}^2$$

Marker index, a product of information content, as measured by PIC, Effective multiplex ratio [5] and Resolving power (Rp) of each primer combination was

Table 1 Name of varieties, their pedigree, releasing centre, cluster and sub-population

S. no.	Name of the varieties	Pedigree	Releasing center	Cluster	Sub-population
1	ADT-1	Selection from Hilly variety	TRRI, Aduthurai	I	C
2	Alankar	D63-6094 X D-61-4249	GBPUAT, Pantnagar	I	C
3	Ankur	SPS from composite of 22 crosses	GBPUAT, Pantnagar	I	C
4	Birsa Soya-1	Spontaneous mutant of Sepaya Black	–	I	C
5	Bragg	Jackson X D 49-2491	GBPUAT, Pantnagar	I	C
6	CO-1	Selection from EC 398321	TNAU, Coimbatore	I	C
7	CO Soya-2	UGM21 X JS 335	TNAU, Coimbatore	I	C
8	Durga	EC14437 X Durga	JNKKV, Sehore, Jabalpur	I	C
9	Gaurav	D60-9647 X EC7034	JNKKV, Sehore, Jabalpur	I	C
10	GS-1	Selection from Punjab 1	–	I	C
11	GS-2	Selection from Geduld variety	–	I	C
12	Hardee	D49-772 X Improved Pelican	MACS, Pune	I	C
13	Hara soya	Himso 1520 × Bragg	HPKV, Palampur	I	C
14	Indirasoya-9	Secondary selection 80–21	IGKV, Raipur	I	C
15	Improved Pelican	Tanloxi X PI 60406	MACS, Pune	I	C
16	JS-2	Selection from Tehari Garhwal material	JNKKV, Sehore, Jabalpur	I	C
17	JS-71-05	Selection from Lee type exotic material	JNKKV, Sehore, Jabalpur	I	D
18	JS-75-46	Improved Pelican X Semmes	JNKKV, Sehore, Jabalpur	I	D
19	JS-76-205	Kalitur X Bragg	JNKKV, Sehore, Jabalpur	IV	E
20	JS-79-81	Bragg X Hara soya-deciduous	JNKKV, Sehore, Jabalpur	IV	E
21	JS-80-21	JS 75-1 X PK 73–94	JNKKV, Sehore, Jabalpur	I	E
22	JS-90-41	PS 73-7 X Hark	JNKKV, Sehore, Jabalpur	IV	E
23	JS-93-05	Secondary selection from PS-73-22	JNKKV, Sehore, Jabalpur	V	E
24	JS-335	JS78-77 X JS 75-5	JNKKV, Sehore, Jabalpur	V	E
25	Kalitur	Indigenous Native Variety	JNKKV, Sehore, Jabalpur	V	E
26	KB-79	Hardee X Monetta	GKVI, Bangalore	V	E
27	KHSB-2	Manloxi X EC 39821	GKVI, Bangalore	V	E
28	Lee	S-100 X CNS	–	V	E
29	LSB-1	Selection from MACS 330	ANGRAU, Hyderabad	V	E
30	MACS-13	Hampton X EC 7034	MACS, Pune	V	E
31	MACS-57	JS 2 X Improved Pelican	MACS, Pune	V	E
32	MACS-58	JS 2 X Improved Pelican	MACS, Pune	V	E
33	MACS-124	JS 2 X Improved Pelican	MACS, Pune	V	E
34	MACS-450	Bragg X MACS 111	MACS, Pune	V	E
35	MAUS-1	Mutant from DS 87-14	MAUS, Parbhani	V	E
36	MAUS-2	Selection from 84 to 14	MAUS, Parbhani	V	E
37	MAUS-32	Selection from JS 80-21	MAUS, Parbhani	V	E
38	MAUS-47	PS 73-7 X Hark	MAUS, Parbhani	V	E
39	MAUS-61	JS 71-1 X PK 73–94	MAUS, Parbhani	V	E
40	MAUS-61-2	JS 80-21 X KB 60	MAUS, Parbhani	II	A
41	MAUS-71	JS 71-05 X JS 87–38	MAUS, Parbhani	II	A
42	MAUS-81	KB74 X JS335	MAUS, Parbhani	II	A
43	Monetta	Exotic X Variety EC 2587	MACS, Pune	II	A
44	NRC-2	Induced Mutant of Bragg	DSR, Indore	II	A
45	NRC-7	Selection from S 69–96	DSR, Indore	II	A
46	NRC-12	Induced Mutant of Bragg	DSR, Indore	II	A
47	NRC-37	Gaurav X Punjab 1	DSR, Indore	II	A
48	Palam Soya	JS72-451 X Punjab1	HPKV, Palampur	II	D

Table 1 continued

S. no.	Name of the varieties	Pedigree	Releasing center	Cluster	Sub-population
49	PK-262	UPSM 97 X Hardee	GBPUAT, Pantnagar	II	D
50	PK-308	T 31 X Hardee	GBPUAT, Pantnagar	II	D
51	PK-327	UPSM 82 X Semmes	GBPUAT, Pantnagar	II	A
52	PK-416	UPSM 534 X S38	GBPUAT, Pantnagar	II	A
53	PK-471	Hardee X Punjab 1	GBPUAT, Pantnagar	II	A
54	PK-472	Hardee X Punjab 1	GBPUAT, Pantnagar	II	D
55	PS-1024	PK 308 X PK 317	GBPUAT, Pantnagar	II	A
56	PS-1029	PK 262 X PK 317	GBPUAT, Pantnagar	II	A
57	PS-1042	Bragg X PK 416	GBPUAT, Pantnagar	II	A
58	PS-1092	PK 327 X PK 416	GBPUAT, Pantnagar	II	A
59	PS-1241	EC 7965 X Bragg	GBPUAT, Pantnagar	II	D
60	Punjab-1	Selection from Nanking Variety	PAU, Ludhiana	III	D
61	Pusa-16	CNS X Lee	IARI, New Delhi	III	D
62	Pusa-22	Punjab 1 X Clark 63	IARI, New Delhi	III	D
63	Pusa-20	Bragg X Lee	IARI, New Delhi	III	B
64	Pusa-24	Shelby X Bragg	IARI, New Delhi	III	B
65	Pusa-37	Bragg X Java 16	IARI, New Delhi	III	D
66	Pusa-40	8-3 X Lee	IARI, New Delhi	III	B
67	RAUS-5	Pusa 16 X JS 335	MPUA&T, Udaipur	III	D
68	Shilajeet	Selection from 9309	GBPUAT, Pantnagar	III	D
69	Shivalik	Selection from Segregating PK 73-55	HPKVV, Palampur	III	B
70	SL-96	Botato X JS 3	PAU, Ludhiana	III	B
71	SL-295	PK 416 X PS 564	PAU, Ludhiana	III	B
72	SL-525	PK 416 X PK 1023	PAU, Ludhiana	III	B
73	Type-49	Selection from Indigenous material	MACS, Pune	III	B
74	VLS-1	Mutant of Bragg	VPKAS, Almora	III	B
75	VLS-2	Selection from VHC 856007	VPKAS, Almora	III	B
76	VLS-21	Selection from VHC 3055	VPKAS, Almora	III	B
77	VLS-47	Selection from KHSF-3-1-1	VPKAS, Almora	III	B
78	Samrat	Farmer's variety	–	III	B

calculated [24]. The Jaccard's similarity index was calculated using NTSYS-pc version 2.02e (Applied Bio-Statistics, Inc., Setauket, NY, USA) package to compute pairwise Jaccard's similarity coefficients [25] and this similarity matrix was used in cluster analysis using an Unweighted Pair-Group Method with Arithmetic averages (UPGMA) to obtain a dendrogram. NTSYS pc version 2.02 [26] was also used to perform Principal Coordinate Analysis (PCoA) to identify multidimensional relationship that describes portions of the genetic variance in the data set. Genetic similarity coefficient was calculated for each pair of cultivars [3] to determine the effectiveness of the group of SSR loci in distinguishing each of the 78 varieties.

Model-based cluster analysis was performed to infer genetic structure and to define the number of clusters (gene pools) in the dataset using the software STRUCTURE version 2.3.4 [27]. The membership of each genotype was

run for range of genetic clusters from value of $K = 2-10$ using a model without admixture and correlated allele frequencies. The authors used the burn-in period of 100,000 and Monte Carlo Markov Chain replicates of 100,000 [28] and the analysis was repeated five times. Ln (PD) was derived for each K and then plotted to find the plateau of the K values [29].

Results and Discussion

SSR Polymorphism

In the present study, 39 STMS markers were used to genotype 78 soybean varieties. The authors have used single locus SSR markers since they provide more reliable scoring of genotypes as compared to multi-locus SSRs. Reproducibility of the amplification patterns was checked

Table 2 Locus, linkage group, chromosome number, range of allele size, no. of alleles per locus, PIC values, major allelic frequency, Dp, MI and Rp

S. no.	Locus	Linkage group	Chromosome number	Range of alleles size	No. of alleles	PIC values	Major allelic frequency	Discrimination power	Marker index	Resolving power
1	Satt211	A ₁	5	150–190	3	0.628	0.438	0.637	1.885	1.999
2	Satt276	A ₁	5	310–400	4	0.685	0.378	0.697	2.743	1.999
3	Satt493	A ₂	8	250–275	4	0.678	0.467	0.687	2.714	1.999
4	Satt415	B ₁	11	300–380	3	0.073	0.962	0.038	0.219	0.151
5	Satt426	B ₁	11	200–230	4	0.653	0.513	0.652	2.613	1.955
6	Satt063	B ₂	14	100–155	7	0.735	0.218	0.851	5.149	1.999
7	Satt126	B ₂	14	125–155	2	0.813	0.703	0.399	1.626	1.081
8	Satt194	C ₁	4	210–250	4	0.401	0.75	0.406	1.602	0.999
9	Satt524	C ₁	4	175	1	0	1	0	0	–
10	Satt100	C ₂	6	140–165	3	0.578	0.461	0.583	0.192	1.998
11	Satt170	C ₂	6	75–90	3	0.184	0.911	0.165	0.554	0.352
12	Sat_114	D ₂	17	260–315	4	0.668	0.426	0.669	2.673	1.998
13	Satt498	D ₂	17	200–250	2	0.025	0.974	0.506	0.051	0.102
14	Satt129	D _{1a}	1	110–150	4	0.672	0.387	0.678	2.688	1.999
15	Satt184	D _{1a}	1	150–180	4	0.597	0.583	0.567	2.389	1.666
16	Satt216	D _{1b}	2	140–220	4	0.632	0.5	0.633	2.531	1.999
17	Satt459	D _{1b}	2	200	1	0	1	0	0	–
18	Satt231	E	15	100–110	2	0.496	0.544	0.502	0.992	1.818
19	Satt411	E	15	100–165	6	0.795	0.359	0.781	4.77	1.999
20	Satt072	F	13	200	1	0	1	0	0	–
21	SOYHSP176	F	13	110–150	7	0.793	0.328	0.806	5.556	1.999
22	Satt038	G	18	155–190	4	0.609	0.545	0.617	2.436	1.817
23	Sat_187	G	18		1	0	1	0	0	–
24	Sat_127	H	12	210–300	6	0.759	0.381	0.771	4.558	1.999
25	Satt434	H	12	310–360	6	0.749	0.351	0.759	4.499	1.999
26	Sat_105	I	20	220–275	4	0.851	0.375	0.721	3.403	1.994
27	Satt587	I	20	125	1	0	1	0	0	–
28	Satt046	J	16	180	1	0	1	0	0	–
29	Satt431	J	16	190–260	6	0.778	0.302	0.797	4.671	1.999
30	Satt539	K	9	175–195	5	0.721	0.407	0.725	3.604	1.999
31	SOYPRP1	K	9	120–150	3	0.338	0.8	0.342	1.015	0.799
32	Sat_099	L	19	210–285	5	0.481	0.417	0.659	2.407	1.999
33	Satt278	L	19	200–280	2	0.264	0.842	0.268	0.529	0.628
34	Satt346	M	7	195–210	2	0.396	0.739	0.391	0.793	1.041
35	Sat_084	N	3	175–195	5	0.931	0.553	0.637	4.659	1.783
36	GMABABE	N	3	155–180	4	0.591	0.536	0.579	2.364	2.086
37	Satt109	O	10	170–300	6	0.773	0.3	0.801	4.639	1.997
38	Satt132	O	10	150–200	3	0.53	0.521	0.52	1.591	1.912
39	GMSC51			150–180	3	0.616	0.421	0.637	1.848	1.999
Average					3.59	0.59	0.54	0.59	2.54	1.64

by performing two independent reactions for each primer pair. The amplification profile of these markers was consistent uniformly and no new differences were detected. A total of 140 alleles were scored from these 39 primer pairs and 84.61% were found polymorphic. The number of alleles per locus varied from 2 to 7, with an average of 3.59

alleles per marker (Table 2). The highest number of alleles was detected at locus Satt063 and SOYSH176 with a total of 7 alleles. The respective values for overall genetic variability for Polymorphism Information Content (PIC), Resolving Power (Rp), Major allelic frequency, Marker index (MI) and Discrimination Power (Dp) across all the

78 genotypes are given in Table 2. PIC values ranged from 0.073 (Satt415) to 0.932 (Satt084), with an average value of 0.59. The present study reported that the genetic diversity index is positively correlated with the number of alleles at a given locus, similar to that earlier reported [30]. The genetic diversity observed in the present study is comparable to those reported in earlier studies [31], 4 alleles per locus were detected with an average PIC value of 0.58. In another study, an average of 4.78 alleles per locus was observed among 92 soybean accessions of Hubei using SSR markers [32]. A slightly higher SSR diversity among soybean accessions was also reported, where 6.91 and 6.3 alleles per locus and an average polymorphic information content of 0.68 and 0.63, respectively, were reported [33, 34]. In another study, a lower SSR diversity was observed, where they detected 3.23 alleles per locus and an average PIC value of 0.38 in 48 vegetable soybean accession using 22 polymorphic EST-SSRs [35]. The possible explanation for this observation is that EST-SSRs are mostly conserved sequences and show lesser variations across genotypes, whereas, in the present study, genomic-SSRs for genetic diversity evaluation were used. MI values ranged from 0.051 (Satt498) to 5.556 (SOYSH176) with an average value of 2.54. In the present study, the average genetic similarity (GS) among the 78 soybean genotypes was found to be 0.41 and most of the values lied between 0.2 and 0.5, reflecting relatively high degree of genetic diversity among the genotypes. The levels of average GS observed was lower as compared to those earlier reported [31], where an average GS value of 0.70 was observed and 42% of the accessions were having values higher than 0.75 among 23 soybean genotypes representing several independent breeding sources from Southeastern Europe and five plant introductions from Western Europe and Canada, using 20 SSR markers. A slightly higher average genetic similarity values were also reported [36], where the average genetic similarities of 0.44 and 0.50 were observed for 40 plant introductions from several Asian and European countries and 39 North American elite cultivars, respectively. On the other hand, an average genetic dissimilarity coefficient of 0.63 and 0.815 was reported for 186 Brazilian soybean cultivars, and for 45 Canadian soybean cultivars and 37 exotic germplasm accessions, respectively [34, 37], which was similar to GS reported in the present study. Similar results based on AFLP were reported earlier [38], where they recorded an average GS of 0.563.

The resolving power (Rp), a feature of marker that indicates the discriminatory potential of the primer, ranged from 0.102 (satt498) to 2.086 (GMABABE) with an average value of 1.64 (Table 2). Rp value takes into account the number of polymorphic bands in a pattern and the informative value of each individual polymorphic band, thus provides a measure of discriminatory power of a

marker. In another study, a lower mean Rp value (1.0) was reported in genetic diversity analysis in emmer wheat using SSR markers [39]. The frequency of major alleles was also calculated which ranged from 0.218 to 1 with an average value of 0.54. Discrimination power (Dp) was calculated for all the polymorphic markers, and varied from 0.038 (Satt415) to 0.851 (Satt063) with an average of 0.59.

Phylogenetic Relationships Among the Soybean Varieties

The genetic relationship among the soybean varieties was assessed by a cluster analysis of the similarity matrix using SSR data. A UPGMA cluster diagram grouped all the 78 soybean varieties into five major clusters, I, II, III, IV and V (Fig. 1) and the clusters were comprised of 19, 20, 19, 3 and 17 genotypes, respectively. The genetic similarity coefficients found in the cultivar comparison matrix were relatively low. The cluster I is further divided into two sub-clusters, and possesses genotypes released from different breeding centers viz., Pantnagar, Jabalpur and Pune, but their parentages were based on US introductions. Cluster II consists of all the varieties of PK, PS and NRC series released from Pantnagar and Indore, respectively and are grouped into different, but same sub-cluster of cluster II. Genotypes such as MAUS61-2, MAUS71 and MAUS81 released from Marathwada Agricultural University (MAUS), Parbhani are also grouped together in cluster II. Similarly, all the genotypes released from New Delhi, Ludhiana and Almora are grouped into different sub-clusters of cluster III, respectively. Cluster IV consists of only the genotypes released from Jabalpur (JS series). All the varieties of MACS series released from Maharashtra Academy for Cultivation of Sciences (MACS), Pune, belong to same sub-cluster of cluster V. Genotypes released from Parbhani viz., MAUS1, MAUS2, MAUS32, MAUS47 and MAUS61 are also clustered together in cluster V. Thus, genotypes released from various parts of India showed diverse clustering patterns. This may be due to the use of diverse parents and breeding material for the development of these varieties. Principal coordinate analysis (PCoA) which is used to show multiple dimensions of the distribution of the genotypes in a scatter-plot (Fig. 2) which was at par with that of grouping as shown in dendrogram (Fig. 1).

The varieties Pusa-16, Pusa-20, Pusa-22, Pusa-40, Pusa-24 and Pusa-37 released from IARI, New Delhi are clustered in the same sub sub-cluster because they have been derived either from Lee or Bragg varieties. Similarly, the varieties PK-262, PK-308, PK-472 derived from Hardee variety and released from Pantnagar are clustered in the same sub sub-cluster of dendrogram. The PS-1024, PS-1029, PS-1024, PS-1092 varieties derived

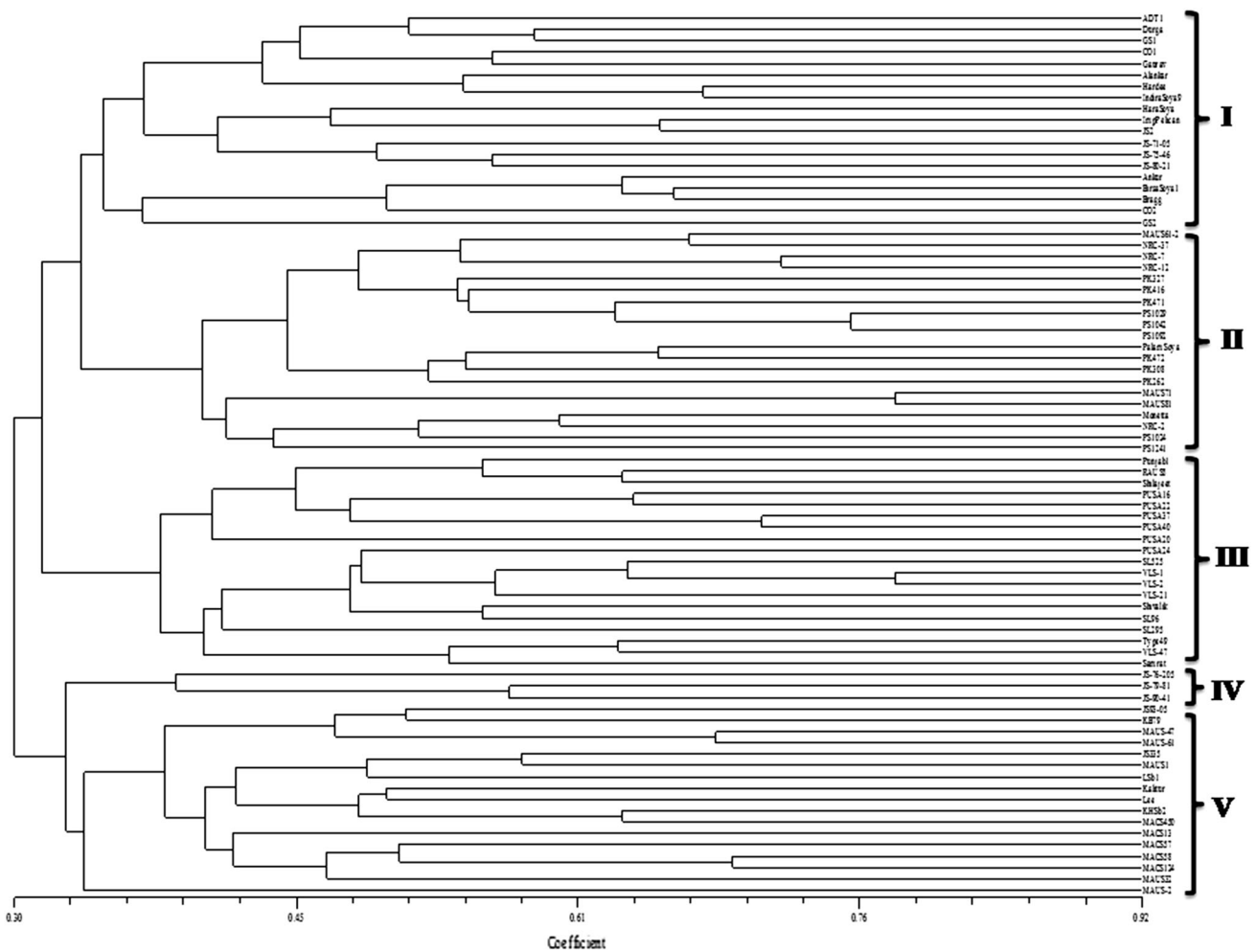


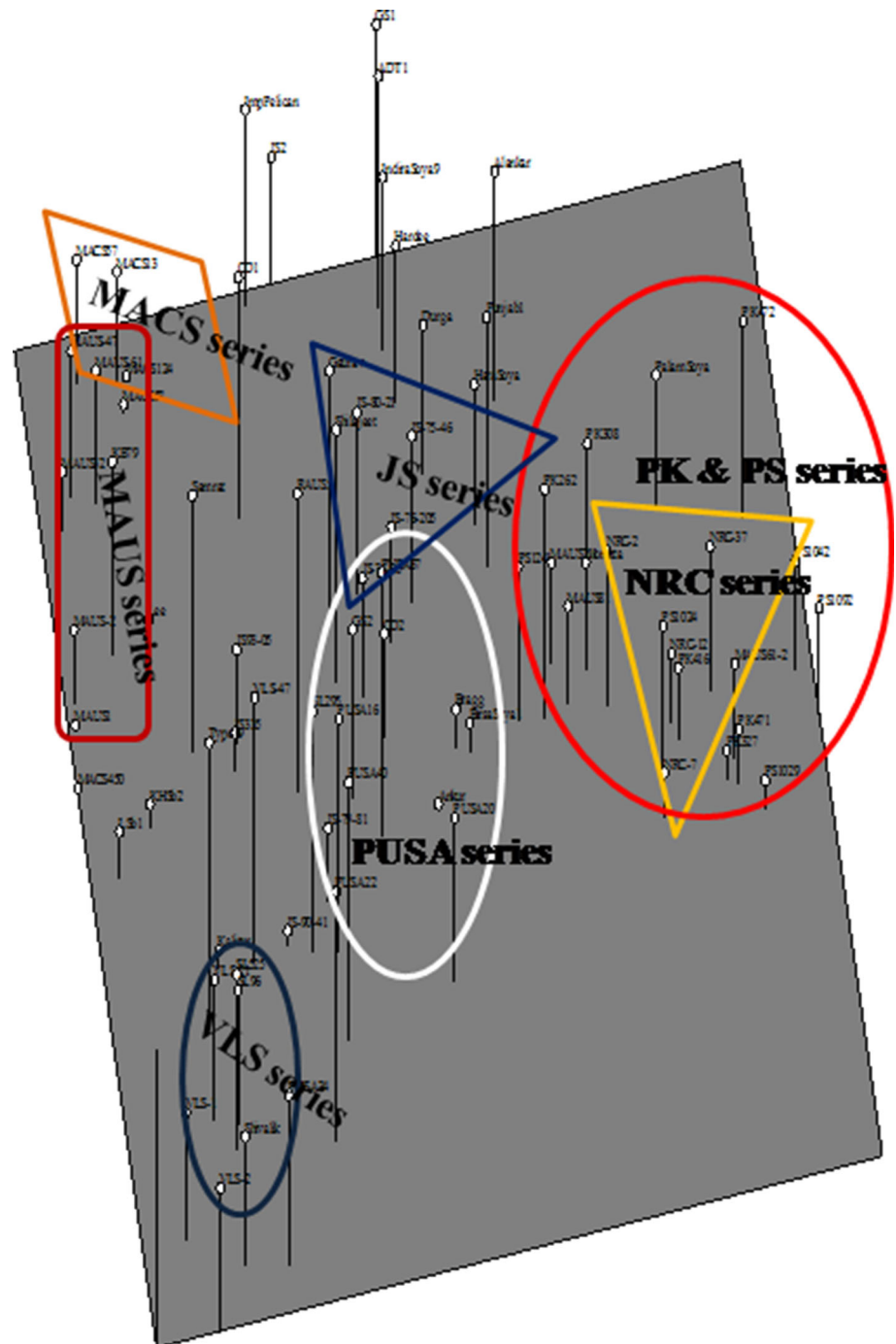
Fig. 1 UPGMA dendrogram showing genetic relationship among soybean varieties based on SSR markers

from either PK-317 or PK416 and released from Pantnagar also fall in the same sub sub-cluster. The varieties MACS-57, MACS-58, MACS-124 and MACS-450 released from Pune, derived from common parents viz., JS-2 and Improved pelican, are clustered into same sub-cluster of cluster V. Similarly, varieties of the MAUS, NRC, VLS, JS and SL series are clustered in same sub sub-cluster of the dendrogram based on their origin, pedigree and releasing center. It is because the concerned breeding centre has used the similar plant material for the breeding of these varieties. The seven SSR markers viz., Satt063, SOYHSP176, Satt411, Sat_127, Satt434, Satt431 and Satt109, which are reported to produce the maximum number of alleles/marker, having high discrimination potential and together are effective to differentiate, discriminate and identify all the soybean varieties used in the present study, thus providing concluding evidence about the use of SSR markers for genetic differentiation and identification of soybean varieties.

Population Structure

The SSR marker data were used for model-based clustering using the software STRUCTURE version 2.3.4 for determining the genetic structure among the 78 soybean varieties. Based on maximum likelihood and delta K (ΔK) values, the number of optimum subgroups was five. The level of genetic differentiation or Wrights fixation index (F_{st}) was calculated through STRUCTURE programme between the five soybean sub-populations assigned to the corresponding A–E and the inferred population structure is given in Fig. 3. The sub-populations (A–E) had an F_{st} value of 0.339, 0.338, 0.251, 0.275 and 0.201, respectively, with an average value of 0.281, indicating moderate population structure. The first inferred population A consists of genotypes released from Indore, Pantnagar and Parbhani. The sub-population B includes varieties belonging to VLS, PUSA and SL series released from Almora, New Delhi and Ludhiana, respectively. The sub-population C consists of varieties released from different breeding centers viz.,

Fig. 2 3-D principal component analysis depicting genetic relationship among 78 soybean varieties



Pantnagar, Jabalpur, Coimbatore and Pune, but their parentages were based on US introductions. Varieties released mostly from New Delhi and Pantnagar belong to sub-population D. Similarly, varieties released from Pune, Parbhani, Bangalore and Jabalpur are grouped into sub-population E. Sub-population A were found to be genetically more differentiated followed by B, D, C and E, based on F_{st} value. The reason lies that varieties belonging to

sub-population A have been developed using different breeding material having diverse origin, whereas the genotypes of sub-population E have originated from similar breeding material released from Jabalpur. Thus, groupings of soybean varieties through STRUCTURE analysis are well in agreement with the distance-based clustering and PCoA, classifying the genotypes on the basis of origin, pedigree and releasing centre.

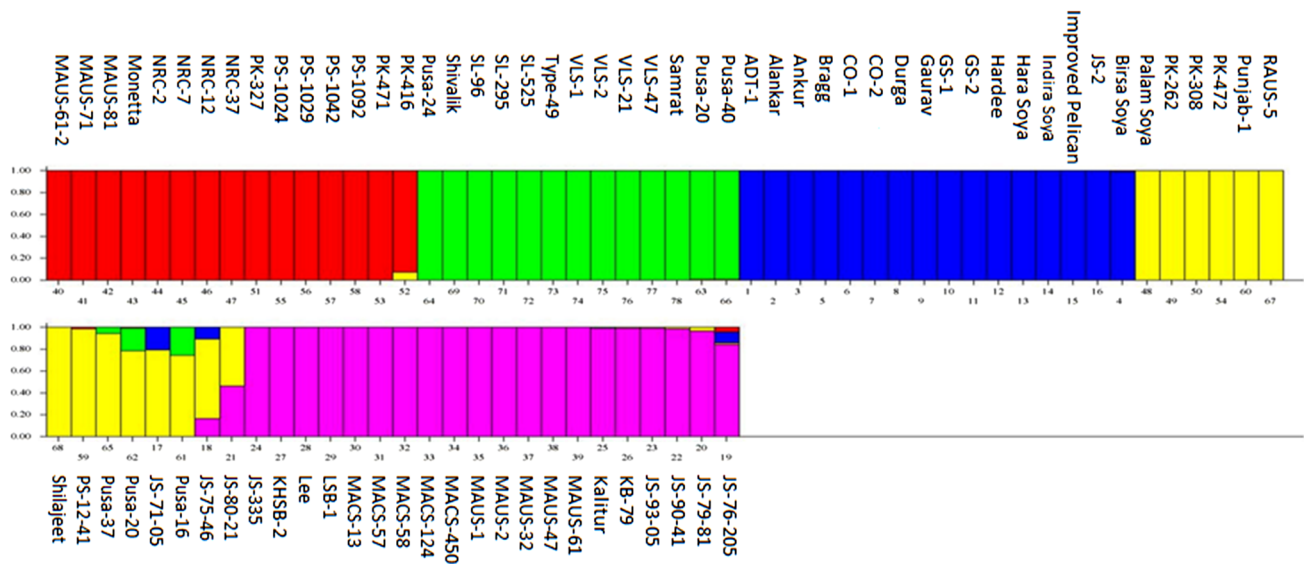


Fig. 3 The population structure of 78 soybean varieties as obtained from STRUCTURE 2.3.4 software

In summary, the results of STRUCTURE and F_{st} analysis were in good agreement with the results obtained through phylogenetic tree-based, similarity coefficient distribution and PCoA analyses, grouping the genotypes on the basis of origin, pedigree and releasing centre and confirmed the presence of statistically high genetic diversity and moderate population structure, which is a prerequisite and basic requirement for carrying out linkage disequilibrium (LD) based association mapping (AM) analysis.

Conclusion

The present study showed that combined analysis with both phylogenetic and population structure analysis is a powerful method for characterizing soybean varieties. Also, highly polymorphic SSR markers identified in the present study can be effectively used for accessing genetic structure, relationship and for further molecular breeding programs of soybean. The present results also showed that the polymorphic SSR loci used in the study were able to differentiate the Indian soybean varieties. Hence, there is a great scope for future improvement of soybean in India by hybridizing distant lines identified during the present study and for further selection of the desired genotypes to derive more agronomically superior varieties for commercial production.

Acknowledgements The authors are thankful to the Director of ICAR-Directorate of Soybean Research (DSR), Indore for providing the lab facilities and Indian Council of Agricultural research (ICAR), New Delhi for providing the financial assistance.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ministry of Agriculture (2014) Statistics at a glance. Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India, New Delhi
2. Struss D, Ahmad R, Southwick SM, Boritzki M (2003) Analysis of sweet cherry (*Prunus avium* L.) cultivars using SSR and AFLP markers. *J Am Soc Hortic Sci* 128:904–909
3. Diwan N, Cregan PB (1997) Automated sizing of fluorescently labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *Theor Appl Genet* 95:723–733
4. Dossett M, Bassil NV, Lewers KS, Finn CE (2012) Genetic diversity in wild and cultivated black raspberry (*Rubus occidentalis* L.) evaluated by simple sequence repeat markers. *Genet Resour Crop Evol* 59(8):1849–1865
5. Powell W, Marchray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1:215–222
6. Abid G, Mingeot D, Udupa SM, Muhovski Y, Watillon B, Sassi K, Mhamdi M, Souissi F, Mannai K, Barhoumi F, Jebara M (2015) Genetic relationship and diversity analysis of faba bean (*Vicia faba* L. var. Minor) genetic resources using morphological and microsatellite molecular markers. *Plant Mol Biol*. doi: [10.1007/s11105-015-0871-0](https://doi.org/10.1007/s11105-015-0871-0)
7. Tavaud-Pirra M, Sartre P, Nelson R, Santoni S, Texier N, Roumet P (2009) Genetic diversity in a soybean collection. *Crop Sci* 49:895–902
8. Yoon M, Lee J, Kim C, Kang J, Cho E, Baek H (2009) DNA profiling and genetic diversity of Korean soybean (*Glycine max* (L.) Merrill) landraces by SSR markers. *Euphytica* 165:69–77
9. Sun B, Fu C, Yang C, Ma Q, Pan D, Nian H (2013) Genetic diversity of wild soybeans from some regions of southern China based on SSR and SRAP markers. *Am J Plant Sci* 4:257–268

10. Velusamy V, Toan VD, Park E (2013) Evaluation of genetic diversity in wild soybean (*Glycine soja* Sieb. and Zucc.) accessions based on SSR markers. *Inter J Agro Plant Prod* 4:1241–1246
11. Ghosh J, Ghosh PD, Choudhury PR (2014) An assessment of genetic relatedness between soybean [*Glycine max* (L.) Merrill] cultivars using SSR markers. *Am J Plant Sci* 5:3089–309612
12. Bizari EH, Uneda-Trevisoli SH, Vianna VF, Meyer AS, Mauro AOD (2014) Genetic diversity in early-maturing soybean genotypes based on biometric and molecular parameters. *J Food Agri Environ* 12:259–265
13. Dong D, Fu X, Yuan F, Chen P, Zhu S, Li B, Yang Q, Yu X, Zhu D (2014) Genetic diversity and population structure of vegetable soybean (*Glycine max* (L.) Merr.) in China as revealed by SSR markers. *Genet Resour Crop Evol* 61:173–183
14. Hu Z, Kan G, Zhang G, Zhang D, Hao D, Yu D (2014) Genetic diversity analysis using simple sequence repeat markers in soybean. *Plant Genet Resour* 12:S87–S90
15. Kumawat G, Singh G, Gireesh C, Shivakumar M, Arya M, Agarwal DK, Husain SM (2015) Molecular characterization and genetic diversity analysis of soybean (*Glycine max* (L.) Merr.) germplasm accessions in India. *Physiol Mol Biol Plants* 21:101–107
16. Torres AR, Grunvald AK, Martins TB, Santos MA, Lemos NG, Silva LAS, Hungria M (2015) Genetic structure and diversity of a soybean germplasm considering biological nitrogen fixation and protein content. *Sci Agric* 72:47–52
17. Bisen A, Khare D, Nair P, Tripathi N (2015) SSR analysis of 38 genotypes of soybean (*Glycine max* (L.) Merr.) genetic diversity in India. *Physiol Mol Bio Plants* 21:109–115
18. Yadav S, Singh A, Singh MR, Goel N, Vinod KK, Mohapatra T, Singh AK (2013) Assessment of genetic diversity in Indian rice germplasm (*Oryza sativa* L.): use of random versus trait-linked microsatellite markers. *J Genet* 92:545–557
19. Arora A, Kundu S, Dilbaghi N, Sharma I, Tiwari R (2014) Population structure and genetic diversity among Indian wheat varieties using microsatellite (SSR) markers. *Aust J Crop Sci* 8:1281–1289
20. Thakur AK, Singh KH, Singh L, Nanjundan J, Rana MK, Singh D (2015) Transferability of SSR markers of *Brassica species* to some popular varieties of *Brassica juncea*. *Proc Natl Acad Sci, India, Sect B Biol Sci* 85:1001–1010
21. Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *BRL Focus* 12:13–15
22. Cregan PB, Jarvik T, Bush AL, Shoemaker RC, Lark LG, Kahler AL (1999) An integrated linkage map of the soybean genome. *Crop Sci* 39:1464–1490
23. Anderson JA, Churchill JE, Autrique SD, Tanksley S, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36:181–188
24. Prevost A, Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet* 98:107–112
25. Jaccard P (1908) Nouvelles recherches sur la distribution florale. *Bull Soc Vaud Sci Nat* 44:223–270
26. Rohlf FJ (2000) NTSYS-pc: Numerical taxonomy and multivariate analysis system, version 2.2. Exeter Software, Setauket, New York
27. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
28. Pritchard JK, Wen W (2003) Documentation for the structure software, version 2. Department of Human Genetics, University of Chicago, Chicago. <http://pritch.bsd.uchicago.edu/software>
29. Evano G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
30. Chotiarnwong O, Chatwachirawong P, Chanprame S, Srinives P (2007) Evaluation of genetic diversity in Thai indigenous and recommended soybean varieties by SSR markers. *Thai J Agri Sci* 40:119–126
31. Ristova D, Sarcevic H, Simon S, Mihajlov L, Pejic I (2010) Genetic diversity in southeast European soybean germplasm revealed by SSR markers. *ACS* 75:21–26
32. Rong Z, Xian-Zhi W, Xiao-Juan Z, Ai-Hua S, Gan-Ying T, Xin-An Z (2007) Genetic diversity analysis by SSR of summer sowing soybean in Hubei. *J Wuhan Bot Res* 25:544–549
33. Xie H, Guan R, Chang R, Qiu L (2005) Genetic diversity of Chinese summer soybean germplasm revealed by SSR markers. *Chin Sci Bull* 50:526–535
34. Fu YB, Peterson GW, Morrison MJ (2007) Genetic diversity of Canadian soybean cultivars and exotic germplasm revealed by simple sequence repeat markers. *Crop Sci* 47:1947–1954
35. Zhang GW, Xu SC, Mao WH, Hu Q, Gong YM (2013) Determination of the genetic diversity of vegetable soybean [*Glycine max* (L.) Merr.] using EST-SSR markers. *J Zhejiang Univ Sci B* 14:279–288
36. 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
37. Priolli RHG, Pinheiro JB, Zucchi MI, Bajay MM, Vello NA (2010) Genetic diversity among Brazilian soybean cultivars based on SSR loci and pedigree data. *Braz Arch Biol Technol* 53:519–531
38. Satyavathi CT, Bhat KV, Bharadwaj C, Tiwari SP, Chaudhury VK (2006) AFLP analysis of genetic diversity in Indian soybean (*Glycine max* (L.) Merr.) varieties. *Genet Resour Crop Evol* 53:1069–1079
39. Salunkhe A, Tamhankar S, Tetali S, Zaharieva M, Bonnett D, Trethowan R, Misra S (2012) Molecular genetic diversity analysis in emmer wheat (*Triticum dicoccon* Schrank) from India. *Genet Resour Crop Evol* 60(1):1–10