RESEARCH ARTICLE

Genetic diversity and population structure of Indian soybean (*Glycine max* (L.) Merr.) as revealed by microsatellite markers

Sharad Tiwari^{1,2} · Niraj Tripathi¹ · Koji Tsuji^{3,4} · Keerti Tantwail¹

Received: 19 December 2018 / Revised: 20 May 2019 / Accepted: 28 May 2019 / Published online: 4 June 2019
© Prof. H.S. Srivastava Foundation for Science and Society 2019

Abstract The genetic structure and diversity of 148 soybean genotypes were analyzed using 26 simple sequence repeat (SSR) markers. SSR analysis showed distinctive polymorphism among the 148 lines. A total of 71 alleles were detected with an average of 2.8 alleles per SSR locus. Polymorphism information content varied from 0.118 to 0.699 with an average of 0.369. These soybean genotypes were divided into 3 subgroups based on STRUCTURE analysis. Further comparison showed that the arithmetic average (UPGMA) and STRUCTURE subgroups in fact were highly coherent, as genotypes in their subsequent classified group exhibited consistency with their origins or pedigree. This analysis provided a deep insight into the genetic structure of soybeans in India and will help us to improve the breeding strategies.

Keywords Genetic diversity · Microsatellites · Soybean · Polymorphism · Population structure

Introduction

Soybean (*Glycine max* (L.) Merrill) belongs to family Fabaceae, subfamily Papilionoideae tribe Phaseoleae and subtribe Glycininea with the genome size of ~ 1100 Mb. Soybean is one of the most important oilseed cash crops, particularly in the state of Madhya Pradesh which contributes 70% of soybean production of India followed by Maharashtra, Rajasthan and Karnataka. Apart from its use as food oil, soybean oil is considered as a future source of fuel and its protein is an important dietary supplement. Diverse uses of soybean make it a more widely desired crop plant with rapidly increasing demand. During the recent past, soybean yield enhancement has been achieved only by 1.3% per year.

Different studies on assessment of genetic diversity in soybean have been carried out using morphological characteristics (Almeida et al. 2011; Cunha et al. 2013) and molecular markers (Tomar et al. 2011; Bisen et al. 2015). Previous reports are very useful in the utilization of soybean resources and establishing breeding strategies. However, most of these works focus on the genetic diversity of soybean or wild soybean germplasm.

In this study, genetic structure and diversity of 148 soybean genotypes of India have been analysed using simple sequence repeat (SSR) markers. These results will assist in gaining insights into the genetic structure of Indian soybean germplasm and thereby planning further breeding strategies.

✉ Sharad Tiwari
shtiwari@gmail.com

¹ Biotechnology Centre, Jawaharlal Nehru Agriculture University, Jabalpur 482004, India

² Present Address: Department of Plant Breeding and Genetics, Jawaharlal Nehru Agriculture University, Jabalpur 482004, India

³ Japan International Cooperation Agency (JICA) Project for Maximization of Soybean Production in Madhya Pradesh, Indore 452001, India

⁴ Present Address: Faculty of Education, Chiba University, Chiba 263-8522, Japan

Table 1 Soybean genotypes and their characteristics used in present analysis

S.	Genotype	Parentage	Recommended regions	Developed by	HC	FC	PH	CH	GT
1	AGS 2	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
2	AGS 16	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
3	AGS 32	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
4	AGS 48	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
5	Bragg	Jackson × D49-22491	Throughout India	Introduction from USA	Green	White	Present	Tawn	S
6	BR-4	–	–	–	Green	White	Present	Tawn	S
7	B 323	–	–	–	Green	White	Present	Tawn	S
8	KDS 321	Sel. From EC 45371	Southern India	MPKV, Rahuri	Violet	Violet	Present	Tawn	S
9	EC 250348	Exotic selection	–	–	Green	White	Present	Tawn	S
10	EC 250608	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
11	EC 251358	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
12	EC 325099	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
13	Himso 1521	–	North India	HPKV, Palampur	Violet	Violet	Present	Grey	S
14	IC 24541	Exotic selection	–	–	Violet	Violet	Present	Tawn	S
15	IC 39220	Exotic selection	–	–	Violet	Violet	Present	Grey	S
16	IC 313230	Exotic selection	–	–	Violet	Violet	Present	Grey	S
17	JS 2	EC 14437 × Bragg	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
18	JS 335	JS 78-77 × JS 71-5	Central India	JNKVV, Jabalpur	Violet	Violet	Absent	Absent	S
19	JS 93-05	Sec. selection from PS 73-22	Central India	JNKVV, Jabalpur	Violet	Violet	Absent	Absent	S
20	JS 95-60	Sec. selection from PS 73-22	Central India	JNKVV, Jabalpur	Violet	Violet	Absent	Absent	D
21	JS 96-31	DS76-1-29 × JS79-277	Central India	JNKVV, Jabalpur	Green	White	Present	Grey	S
22	JS 97-51	Sel. from EC 34117	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
23	JS 97-52	PK 327 × L129	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
24	JS 97-55	Sel. from EC 241811	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Grey	I
25	JS 98-61	Moretta × Bragg	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	I
26	JS 98-66	DSbl × AGS 79	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	I
27	JS 98-67	DSbl × AGS 79	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	I
28	JS 99-80	DSbl × AGS 79	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	I
29	JS 99 -88	JS 94-67 × JS 90-41	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	I
30	JS 20-1	JS335 × EC 109541	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
31	JS 20-05	JS 94-65 × JS 94-74	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
32	JS 20-06	DSbl × SL 603	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
33	JS 20-09	JSM 246 × NRC 37300	Central India	JNKVV, Jabalpur	Violet	Violet	Absent	Grey	S
34	JS 20-14	JS94-65 × JS GY 94-67	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
35	JS 20-15	JS 335 × PK 1308	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	S
36	JS 20-16	JS94-67 × JS 90-41	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	I
37	JS 20-18	JS 98-61 × SL 659	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
38	JS 20-19	MACS 330 × L129	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
39	JS 20-20	JS 98-61 × EC 333922	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
40	JS 20-21	JS 335 × SL 637	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
41	JS 20-23	JS98-61X SL 633	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
42	JS 20-25	JS 98-62 × SL 633	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
43	JS 20-29	JS 97-52 × JS 95-56	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
44	JS 20-32	JS 98-63 × JS 93-05	Central India	JNKVV, Jabalpur	Violet	Violet	Absent	Absent	D
45	JS 20-34	JS98-63 × PK 768	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	D
46	JS 20-41	JS 97-52 × JS 20-02	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	S
47	JS 20-50	JS 97-52 × JSM 286	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
48	JS 20-53	JS 97-52 × JS 20-02	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S

Table 1 continued

S.	Genotype	Parentage	Recommended regions	Developed by	HC	FC	PH	CH	GT
49	JS 20 -59	JS 97-52 × JSM 52	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	D
50	JS 20-62	JS 97-52 × JS 95-56-B ₁	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
51	JS 20-63	JS 97-52 × JS 95-56-B ₂	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
52	JS 20-64	JS 97-52 × JS(IS)90-5-12-1	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
53	JS 20-65	JS 97-52 × JS(IS)90-5-12-1	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
54	JS 20-68	JSM 240 × JSM 189	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	D
55	JS 20-69	JS 97-52 × SL710	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
56	JS 20-70	JS 97-52 × JSM 120A	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	D
57	JS 20-71	JS 97-52 × JS(15)90-5-12-1	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
58	JS 20-72	JS 99-81 × JS 99-83	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
59	JS 20-73	JS 97-52 × JS 95-56	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
60	JS 20-74	JS 97-52 × SL-710	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
61	JS 20-75	JS 97-52 × JSM 52	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	S
62	JS 20-76	JS 97-52 × SL 710	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
63	JS 20-77	JS 98-67 × SL 633	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	I
64	JS 20-78	JS 98-61 × EC-333922	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
65	JS 20-79	JS 97-52 × JS(15)90-5-12-1	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
66	JS 20-80	JS 97-52 × JSM 52	Central India	JNKVV, Jabalpur	Violet	Violet	Absent	Absent	S
67	JS 20-81	JS 98-62 × SL 633	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
68	JS 20-82	JS 97-51 × EC251348	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
69	JS 20-83	JS95-60 × JS-83	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	D
70	JS 20-84	JS 98-63 × PK 768	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	D
71	JS 20-85	JS 335 × PK 1308	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
72	JS 20-86	JS20-04 × Brazil	Central India	JNKVV, Jabalpur	Green	White	Absent	Absent	S
73	JS 20-87	JS 9752 × JS(15)90-5-12-1	Central India	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
74	JS 20-88	JS 97-52 × JSM 299	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	I
75	JS 20-89	JS 97-52 × JSM 286	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
76	JS 20-90	JS 97-52 × JS 95-56	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	D
77	JS 20-91	JSM 146 × JSM 1525	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
78	JS(SH) 200-3	JS (SH) 88-66 × JS 335	Central India	JNKVV, Jabalpur	Green	White	Present	Tawn	S
79	MACS 1188	JS (SH) 93-01 × MACS 13	Southern India	ARI, Pune	Green	White	Present	Tawn	S
80	NRC 2	Mutant of Bragg	Central India	DSR, Indore	Green	White	Present	Tawn	S
81	NRC 7	Sel. From S-69-96	Central India	DSR, Indore	Violet	Violet	Present	Tawn	S
82	NRC 12	Mutant of Bragg	Central India	DSR, Indore	Violet	Violet	Present	Tawn	S
83	NRC 37	Gaurav × Pb-1	Central India	DSR, Indore	Green	White	Present	Tawn	S
84	NRC 57	JS 335 × DS 369	Central India	DSR, Indore	Violet	Violet	Present	Tawn	S
85	NRC 86	RKS 15 × EC 481309	Central India	DSR, Indore	Violet	Violet	Present	Tawn	S
86	NRC 2320	NA	Central India	DSR, Indore	Violet	Violet	Present	Tawn	S
87	P-955	NA	North India	–	Violet	Violet	Present	Grey	S
88	PB-1	Sel. From Nanking variety	North India	–	Violet	Violet	Present	Tawn	D
89	PS 1225	PK 515 × PK 327	North India	GBPU A&T, Pantnagar	Green	White	Present	Grey	S
90	PS 673	NA	North India	GBPU A&T, Pantnagar	Violet	Violet	Present	Tawn	S
91	PS 1308	T 49 × PS 1024	North India	GBPU A&T, Pantnagar	Green	White	Present	Tawn	D
92	PS 73-22	D-69-3881	North India	GBPU A&T, Pantnagar	Violet	Violet	Present	Grey	D

Table 1 continued

S.	Genotype	Parentage	Recommended regions	Developed by	HC	FC	PH	CH	GT
93	PLSO 79	–	–	–	Violet	Violet	Present	Tawn	D
94	PLSO 90	–	–	–	Violet	Violet	Present	Grey	D
95	RKS-24	PK 472 × PK 1024	Central India	ZARS, Kota, Rajasthan	Green	White	Present	Tawn	S
96	SL 96	Botato × JS-3	North India	PAU, Ludhiana	Green	White	Present	Tawn	D
97	SL 525	PK 416 × PK 1023	North India	PAU, Ludhiana	Green	White	Present	Tawn	D
98	SL 599	PK 416 × SL 269	North India	PAU, Ludhiana	Green	White	Present	Tawn	S
99	SL 738	SL 317 × E 4	North India	PAU, Ludhiana	Green	White	Present	Tawn	D
100	SL 742	SL 457 × SL 459	North India	PAU, Ludhiana	Green	White	Present	Tawn	D
101	SL 747	SL 517 × SL 295	North India	PAU, Ludhiana	Green	White	Present	Tawn	D
102	SISTA 194	–	–	NA	Violet	Violet	Present	Tawn	S
103	VLS-58	DS 54 × VLS 2	North India	VPKAS, Almora	Green	White	Present	Tawn	S
104	VLS-73	PRS 8901 × PK 564	North India	VPKAS, Almora	Violet	Violet	Absent	Absent	D
105	WT-89	–	–	NA	Violet	Violet	Present	Tawn	D
106	JSM 184	JS 335 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
107	JSM 119	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
108	JSM 120	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
109	JSM 122	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Absent	Absent	S
110	JSM 185	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	D
111	JSM 188	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
112	JSM 189	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
113	JSM191	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	I
114	JSM 195	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
115	JSM 265	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
116	JSM 266	JS 335 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
117	JSM -30	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Absent	Absent	D
118	JSM -131	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
119	JSM 139	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
120	JSM 203	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
121	JSM 207	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Absent	Absent	D
122	JSM 269	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
123	JSM 271	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	D
124	JSM 275	JS 93-05 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
125	JSM-104	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
126	JSM-154	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
127	JSM-155	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
128	JSM-220	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	I
129	JSM-223	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	I
130	JSM-224	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
131	JSM-225	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
132	JSM-226	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
133	JSM-227	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
134	JSM-228	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
135	JSM-236	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
136	JSM-276	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
137	JSM-277	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	D
138	JSM-284	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Absent	Absent	S
139	JSM-298	NRC 37 (250 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
140	JSM-115	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	D

Table 1 continued

S.	Genotype	Parentage	Recommended regions	Developed by	HC	FC	PH	CH	GT
141	JSM-155	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D
142	JSM-171	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Absent	Absent	S
143	JSM-239	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
144	JSM-250	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
145	JSM-258	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	S
146	JSM-285	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
147	JSM-294	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Green	White	Present	Tawn	S
148	JSM-306	NRC 37 (300 GY)	Mutant line	JNKVV, Jabalpur	Violet	Violet	Present	Tawn	D

HC hypocotyl colour, FC flower colour, PH presence of hairs, CH colour of hairs, GT growth type (D determinate, I indeterminate, S semideterminate), S. serial number

Materials and methods

Plant materials

Genotypes under study include olden days and present-day cultivars as well as promising fixed mutant lines of some very popular varieties namely JS 335, JS 93-05 and NRC 37. Genotype identity, parentage and different morphological traits such as hypocotyl colour, flower colour, presence of hairs, color of hairs and growth type for five plants of each genotype were used for characterization of 148 soybean genotypes listed in Table 1.

SSR analysis

Genomic DNA was isolated from pooled young leaves of five plants of each genotype selected as representative plants for morphological analysis using CTAB protocol (Saghai Maroof et al. 1984) with necessary modifications. The quality of DNA was checked by horizontal submarine gel electrophoresis on 0.8% agarose gel along with the λ Hind-III standard DNA. Fifty four SSRs markers evenly distributed on 20 soybean genetic linkage groups (Cregan et al. 1999) were firstly screened for polymorphism among some genotypes. Twenty six of them (Table 2) with clearly amplified polymorphic bands were further used in the genotyping of all the genotypes. Information of these SSR primers is available at <http://www.soybase.org>.

The PCR reaction was conducted in a reaction volume of 10 μ l containing 1 $\times$ PCR buffer, 100 μ M dNTPs, 0.4 μ M of each primer, 1.2 mM MgCl₂, 0.5 units *Taq* DNA polymerase (recombinant) and 25 ng template DNA. PCR amplification was performed with initial denaturation at 94 °C for 5 min followed by 35 cycles of 94 °C for 30 s,

45–65 °C for 30 s, 72 °C for 45 s and final extension at 72 °C for 5 min before cooling down to 4 °C.

The PCR products for SSR were resolved by electrophoresis on 3% (w/v) agarose gels stained with ethidium bromide (10 mg/ml) and photographed under UV-light in Syngene Multigenious Bioimager. Molecular size of bands was estimated using a wide range 100 bp ladder (Fermentas Life Sciences, USA).

Analyses of genetic diversity and population structure

Amplified fragments of different sizes were considered as different alleles. DNA bands that were amplified by a given primer were scored as present (1) or absent (0) for all the samples under study. In order to determine the utility of these markers, number of amplicons/alleles per marker, major allele frequency, polymorphic information content (PIC), effective multiplex ratio (EMR)/resolving power (RP), discrimination power (DP) and marker index (MI) were calculated. The PIC value of individual primer was calculated by the formula $PIC = 1 - \sum_{i=1}^n P_{ij}^2$ (Jain et al. 2004). MI, a product of information content, as measured by PIC and EMR was calculated (Powell et al. 1996). RP and DP of each primer combination were calculated using standard methods proposed by Anderson et al. (1993) and Prevost and Wilkinson (1999) respectively. Dendrogram was generated using software PowerMarker version 3.25 (Liu and Muse 2005) to estimate the unweighted pair group method with arithmetic average (UPGMA) and the Nei's genetic distances (Nei 1972) between genotypes.

Population structure of these 148 genotypes was estimated using STRUCTURE V2.3.4 software (Pritchard et al. 2000). Admixture and independent allele frequencies

Table 2 Details of SSR markers used for diversity analysis among 148 soybean genotypes and their characterization

S.	Primer	Primer sequence 5'–3'	Trait associated	LG	Motif	References
1.	Satt119	F: TGTGCCAGTGTGATAGTTA R: CTGATCCCCAATAAATCTG	Seed fucose 1-1	A2	(ATT) ₁₃	Cregan et al. (1999)
2.	Satt245	F: AACGGGAGTAGGACATTTTATT R: GCGCCTCTGAATTTCAAAGAATGAAGA	Pubescence density	M	(ATT) ₁₃ (ATG)	Cregan et al. (1999)
3.	Satt281	F: AAGCTCCACATGCAGTTCAAAAC R: TGCATGGCACGAGAAAGAAGTA	Seed yield	C2	(ATT) ₁₉	Cregan et al. (1999)
4.	Sat_246	F: GCGCATGGTTTACAGATTACTTTATTTTCTA R: GCGGCAATCATTAAATTTATAATGATATAT	Seed weight	C2	(AT) ₂₂ (GT)	Cregan et al. (1999)
5.	Satt433	F: ATCAGCTGCATAAGTAAAGATACA R: ACATGCTTGCAGAGGTGAATA	Photoperiod insensitivity	C2	(ATT) ₁₆	Cregan et al. (1999)
6.	Satt177	F: CGTTTCATTCCCATGCCAATA R: CCCGCATCTTTTCAACCAC	Seed weight	A2	(ATT) ₁₆	Cregan et al. (1999)
7.	Satt527	F: GCGGTACATCTTGCAAATAAATTAAC R: GCGGAATTTGCACATAAATTAATAACT	Plant height, seed weight	L	(ATT) ₁₇	Cregan et al. (1999)
8.	Satt448	F: GCGCTAAGGGCAATTTTATTCAA R: GCGCAGCCTGTTTCTTTCTTTGTC	Phytoph 9-6	L	(ATT) ₁₅	Cregan et al. (1999)
9.	Satt160	F: TCCACACAGTTTTTCATATAATATA R: CATCAAAAGTTTATAACGTGTAGAT	Hypocotyl length, seed coat cracking	F	(ATT) ₃₀	Cregan et al. (1999)
10.	Sat_228	F: GCGTGACTACGGGAAGTTGGAAC R: GCGTTGGCGGTAAGAGCACTATA	Seed palmitic	F	(AT) ₂₄	Song et al. (2004)
11.	Satt686	F: ACGGAAAATAAATGAACTAAGA R: GCGCTATCAGATAGAGAAGCAGAAGAAT	Flowering time	J	(ATT) ₁₄	Cregan et al. (1999)
12.	Satt184	F: GCGCTATGTAGATTATCCAAATTACGC R: GCCACTTACTGTTACTCAT	Flowering time, seed weight	D1a	(ATT) ₁₃	Cregan et al. (1999)
13.	Satt236	F: GCGTGCTTCAAACCAACAACTTA R: GCGGTTTGCAGTACGTACCTAAAATAGA	Flowering time	G	(ATT) ₁₉	Cregan et al. (1999)
14.	Satt235	F: GCGGGCTTTGCCAAGAAGTTT R: GCGGTGAGGCTGGCTATAAG	Plant height	G	(ATT) ₁₆	Cregan et al. (1999)
15.	Satt418	F: GCGAAAGCACATATGGGTTTGAAT R: GCGAGGGCATATATATGATGAGGTA	Plant height	L	(ATT) ₂₇	Cregan et al. (1999)
16.	Satt518	F: GCGCATATCAAATTGCATATAAAATACG R: GCGGGAATATAAAATAAAAATGCTCACTT	Plant height	D2	(ATT) ₁₉	Cregan et al. (1999)
17.	Satt154	F: AGATACTAACAAGAGGCATAAACT R: AAAGAAACGGAATAATACTACATT	100 Seed weight	D2	(ATT) ₂₀	Cregan et al. (1999)
18.	Satt442	F: CCTGGACTTGTTTGCTCATCAA R: GCGGTTCAAGGCTTCAAGTAGTCAC	100 Seed weight	H	(ATT) ₃₅	Cregan et al. (1999)
19.	Satt042	F: GACTTAATTGCTTGCTATGA R: GTGGTGCACTCACTT	Number of nodes	A1	(ATT) ₂₇	Cregan et al. (1999)
20.	Satt130	F: TAAACGAAATTTAGTTTTAAGACT R: TGAATGGCTAAAAACGTGATT	Maximum internode length	G	(ATT) ₁₄	Cregan et al. (1999)
21.	Satt292	F: GCGGAATTAGAACTCCAGTAAAGA R: GCGAGGCCAACATTGAAAAGT	100 Seed weight, Pod maturity	D1b	(ATT) ₁₆	Cregan et al. (1999)
22.	Satt280	F: GCGGAATCTGCTTATTCATTGTGTG R: GCGCCATGCTGTAACACGTCAAT	Days to maturity	J	(ATT) ₁₄	Cregan et al. (1999)
23.	Satt275	F: GCGGGATAATTGGTTTTACGAAAATGC R: GCGCCTAATCACCTAAAAAACGTTTA	Plant height	G	(ATT) ₁₁	Chang et al. (2011)

Table 2 continued

S.	Primer	Primer sequence 5'–3'	Trait associated	LG	Motif	References
24.	Satt114	F: GGGTTATCCTCCCAATA R: ATATGGGATGATAAGGTGAAA	100 Seed weight	F	(ATT) ₁₇	Cregan et al. (1999)
25.	Satt510	F: GCGAGTTTCGCCGTTACCACCTCAGCTT R: CCCTCTTATTTACCCCTAAGACCTACAA	100 Seed weight, seed protein	F	(ATT) ₂₁	Cregan et al. (1999)
26.	Sat_076	F: GCGTAATTAACACCAATATATGACATG R: GCGGGGTAAAAATTCAAAATGT	Seed coat cracking	C2	(AT) ₄₀	Cregan et al. (1999)

S. serial number, LG linkage group

Fig. 1 UPGMA dendrogram showing relationship among 148 soybean genotypes based on morphological data. Colours indicate groups and subgroups of soybean genotypes

model was used in estimating proper subgroups. The number of subgroups (K) was set from 1 to 10 with 10 runs performed separately. In the present analysis, the burn-in period of 50,000 and Monte Carlo Markov Chain replicates of 100,000 and allele frequencies in correlation was used (Pritchard and Wen 2003). The maximum likelihood run

was applied to allocate individual genotypes into a specific group. Further, the genotypes were assigned their groups according to affiliation probabilities (inferred ancestry). The genotypes with 80% affiliation probabilities or more were grouped to a separate group in comparison to those with < 80%. According to grouping, the genotypes

Fig. 2 Electrophoretic illustration of marker Satt527 showing banding pattern with 148 soybean genotypes

with < 80% affiliation probability were considered as admixture which means the genotypes give the impression to have a mixed parentage belonging to dissimilar gene pools or sources. The importance of population delineation clustered by STRUCTURE 2.3.4 was additionally investigated by performing a testing of molecular variance (AMOVA) with Arlequin 3.5 (Excoffier and Lischer 2010). The F_{ST} values were calculated using Arlequin software.

Results and discussion

Availability of genetic diversity within the crop species provides a base for its improvement. The cultivated soybean varieties are the results of the selection from diverse sources available in various environments. Soybean improvement involves recognition of vastly dissimilar germplasm with decidedly polymorphic molecular markers which are important for the mapping of genes/QTLs for significant traits and their consequent use in molecular breeding. Identification of dissimilar genotypes with molecular markers is helpful over the conventional breeding. SSRs molecular markers find its application in analysis of genetic diversity, identification of genotypes and population structure assessment in soybean (Bisen et al. 2015; Torres et al. 2015).

Diversity in the soybean collection

All the 148 soybean accessions were categorized on the basis of their morphological characteristics and UPGMA clustering grouped them accordingly (Fig. 1). In addition, they were genotyped with 26 traits linked microsatellite markers (Table 2) and were categorized on the basis of their ability to generate amplicons, polymorphism level and reliability of the pattern. Out of 26 traits linked microsatellite markers, all were found to be polymorphic revealing more than one allele in all the genotypes. As an example, gel picture showing banding pattern of 148 genotypes of soybean with Satt527 marker is given in Fig. 2. All the amplified 71 allele were found to be polymorphic (100%) with an average of 2.73 polymorphic alleles per primer (Table 2). Similar results have been observed by different groups working on soybean genetic diversity based on microsatellite markers as 53 soybean cultivars with 53 SSR markers amplified alleles with an average of 2.34 per locus (Vieira et al. 2009). In one of our earlier studies, 94% polymorphic alleles with an average of 2.10 per primer were obtained (Tomar et al. 2011). In another study, 97% polymorphic alleles with an average of 3.21 have been reported (Bommi and Ferguson 2005) while 95% alleles amplified by SSR markers were found to be

Table 3 List of microsatellite markers used for genotyping soybean genotypes along with their genetic diversity parameters

S.	Primer	NA	MAF	PIC	MI	RP	DP
1.	Satt119	3	0.81	0.26	0.70	0.76	0.38
2.	Satt245	5	0.49	0.69	1.49	0.96	0.68
3.	Satt281	4	0.42	0.55	1.28	0.82	0.44
4.	Sat_246	3	0.52	0.32	0.98	0.78	0.42
5.	Satt433	3	0.92	0.12	0.38	0.83	0.54
6.	Satt177	3	0.73	0.31	0.71	0.46	0.29
7.	Satt527	3	0.99	0.01	0.30	0.65	0.36
8.	Satt448	3	0.58	0.45	1.03	0.71	0.55
9.	Satt160	4	0.51	0.54	1.39	0.88	0.51
10.	Sat_228	2	0.73	0.31	0.33	0.39	0.19
11.	Satt686	3	0.81	0.25	0.68	0.77	0.23
12.	Satt184	4	0.39	0.68	1.29	0.84	0.39
13.	Satt236	4	0.42	0.55	1.26	0.79	0.41
14.	Satt235	3	0.52	0.52	0.97	0.44	0.42
15.	Satt418	3	0.92	0.22	0.46	0.49	0.39
16.	Satt518	2	0.73	0.31	0.31	0.38	0.20
17.	Satt154	3	0.99	0.21	0.44	0.68	0.37
18.	Satt442	4	0.58	0.45	0.89	0.95	0.56
19.	Satt042	4	0.51	0.54	1.07	0.92	0.49
20.	Satt130	3	0.73	0.31	0.88	0.83	0.41
21.	Satt292	3	0.89	0.26	0.69	0.73	0.33
22.	Satt280	2	0.73	0.32	0.31	0.66	0.28
23.	Satt275	3	0.93	0.21	0.38	0.87	0.42
24.	Satt114	4	0.55	0.46	0.99	0.91	0.51
25.	Satt510	4	0.50	0.55	1.37	0.72	0.44
26.	Sat_076	2	0.67	0.34	0.31	0.56	0.29
Total		71	18.78	9.74	20.89	17.57	10.5
Average		2.73	0.722	0.375	0.803	0.676	0.404

NA number of alleles, *MAF* major allele frequency, *PIC* polymorphism information content, *MI* marker index, *RP* resolving power, *DP* discrimination power

polymorphic among 40 soybean genotypes (Narvel et al. 2000).

Genetic diversity for specific locus was evaluated by polymorphic information content (PIC). PIC value is an indication of allelic diversity and frequency among the genotypes analyzed and in the present study, eight markers exhibited PIC values greater than 0.5. Highest PIC value (0.69) was observed for the primer Satt245 and lowest PIC value (0.01) was recorded for the primer Satt527 (Table 3) with an average of 0.375. However, in an earlier study, PIC ranging from 0.7447 (Satt423) to 0.8585 (Satt155) with an average of 0.8040 was obtained from 178 accessions of Korean soybean using 9 microsatellite markers (Velusamy et al. 2013). In our study, all the primers were polymorphic.

This high rate of SSR polymorphism may be attributed to the selected set of SSR markers which were already tested for polymorphism among a set of genotypes (Tomar et al. 2011). The highest genetic diversity in the present study among soybean genotypes supports previous statement (Powell et al. 1996) that SSRs have elevated discriminatory power compared to other molecular markers.

In genotypic characterization, UPGMA cluster analysis grouped soybean genotypes initially into two groups (Fig. 3). Group A contained ten mutant soybean lines derived from NRC 37 and group B had one hundred thirty eight genotypes with some mutant lines derived from JS 335 and JS 93-05. Group A further divided into two subgroups A1 and A2. Subgroup A1 had only two genotypes while A2 had eight genotypes. Group B was further divided into two subgroups B1 and B2. Subgroup B1 contained twenty seven genotypes including the genotypes of exotic selection, while B2 had 111 remaining genotypes. Group B contained all the varieties released by Jawaharlal Nehru Agriculture University, Jabalpur, India and Directorate of Soybean Research, ICAR, Indore, India. This similarity among group B can be attributed to the use of common ancestors for the development of these genotypes. Subgroup B2 was further divided into C and D. Subgroup C contained sixty one genotypes while subgroup D had fifty genotypes. Subgroup C had two parts C1 and C2. Part C1 contained only nine genotypes and part C2 had fifty two genotypes. Among a total of 43 mutant lines, 33 were grouped in part C2. Subgroup D was divided into D1 and D2. Part D1 contained forty one genotypes while part D2 had only nine genotypes.

The relevant values for genetic variation across all the 148 genotypes obtained after application of trait linked markers in the form of polymorphism information content, resolving power, major allele frequency, discrimination power and marker index are given in Table 3. The MI values ranged from 0.30 to 1.49 with an average of 0.84. The RP is an attribute of marker that represents the discriminatory potential of the primer and it ranged from 0.38 to 0.96 with an average of 0.72 for polymorphic marker. Major allele frequency ranged from 0.39 to 0.99 among polymorphic markers with an average of 0.68 (Table 3 and Fig. 4). Whereas, the DP values ranged from 0.19 to 0.68 with an average of 0.40. The allele number per locus varied from 2 to 5 with an average of 2.73 alleles per locus (Table 3).

Fig. 3 UPGMA dendrogram showing relationship among 148 soybean genotypes based on SSR data

Fig. 4 Major allele frequency of polymorphic SSR markers used for population STRUCTURE analysis of 148 soybean genotypes

Population structure

Admixture model based structure simulations were performed by shifting K from 1 to 10 with 10 iterations using all the 148 genotypes and 26 polymorphic markers. The entire population was divided into three groups A, B and C indicating 33.78% (50), 23.65% (35), and 42.57% (63) of genotype correspondingly and the inferred population structure is given in Fig. 5. Availability of genetic dissimilarity in groups was tested using F_{st} statistics. The groups (A, B and C) showed F_{st} values of 0.7626, 0.7809 and 0.9735 respectively with an average of 0.8398 representing high population structure. Therefore, most structured group was C followed by B and A. The group A and B consisted genotypes of mixed origins however, group C consisted most of the mutant lines developed at Jawaharlal Nehru Agriculture University, Jabalpur.

Fig. 5 Assignment of 148 soybean into three subpopulations (a, b and c) using STRUCTURE 2.3.4 software

Table 4 Analysis of molecular variance (AMOVA)

Source of variation	<i>d.f.</i>	Sum of squares	Variance components	Percentage variation
Among groups	2	148.308	1.17324	31.39
Within groups	148	522.366	1.78527	69.61
Total	150	522.366	2.95851	

Three groups generated from structural analysis in the present study were also subjected to analysis of variance (AMOVA) to assess the percentage of diversity among populations and within population. In the total genetic variance among group, 31.39% was accredited to the populations on the basis of structure, and the remaining 69.61% was due to divergence among the individual within a group (Table 4). The results of AMOVA and F_{st} analysis confirm the results obtained through cluster analysis, similarity coefficient distribution and structure analysis indicating the presence of statistically moderate genetic diversity with high population structure. The results obtained through structure analysis and distance-based clustering are in agreement with each other except few of the mutant genotypes. This confirmation is a decisive factor before carrying out association mapping (AM) analysis.

Conclusion

Results of the present study indicated the studied Indian soybean germplasm has moderate diverse genetic base and high population structure. Hence, the highly diverse

genotypes found in this study can be utilized for the future soybean breeding programmes. Also, the genetically variable genotypes and highly polymorphic SSR markers identified throughout this study can be used for the association mapping (AM), QTLs/genes identification for different quality traits of soybean.

Acknowledgements Research was conducted in “Technical Cooperation Project for Maximization of Soybean Production in Madhya Pradesh” supported by Japan International Cooperation Agency (JICA), Japan.

Compliance with ethical standards

Conflict of interest Authors declare that they have no conflicts of interests.

References

- Almeida RD, Peluzio JM, Afferri FS (2011) Divergência genética entre cultivares de soja, sob condições de várzea. *Rev Ciênc Agron* 42:108–115
- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36:181–186

- 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 Biol Plants* 21:109–115
- Bommi P, Ferguson DL (2005) Soybean cultivars Identification within a selected group using only an agarose gel system with simple sequence repeats DNA markers. *Soybean Genet Newsl* 32:1–7
- Chang SH, Teng TT, Ismail N (2011) Screening of factors influencing Cu (II) extraction by soybean oil-based organic solvents using fractional factorial design. *J Environ Manage* 92(10):2580–2585
- 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. *Crop Sci* 39:1464–1490
- Cunha MCG, Hamawaki OT, Sousa LB (2013) Genetic variability among 79 soybean progenies from UFU breeding program. *Biosci J* 29:340–349
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567
- Jain S, Jain RK, McCouch SR (2004) Genetic analysis of Indian aromatic and quality rice (*Oryza sativa*) germplasm using panels of fluorescently-labeled microsatellite markers. *Theor Appl Genet* 109:965–977
- Liu K, Muse S (2005) Power marker: new genetic data analysis software. Release 3.0. <http://powermarker.net>; verified 9 September
- Narvel JM, Fehr WR, Chu WC, Grant D, Shoemaker RC (2000) Simple sequence repeats diversity among soybean plant introductions and elite genotypes. *Science* 40:1452–1458
- Nei M (1972) Genetic distance between populations. *Am Nat* 106:283–292
- Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1:215–222
- 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
- Pritchard JK, Wen W (2003) Documentation for STRUCTURE software: Version 2. Available from <http://pritch.bsd.uchicago.edu>
- Pritchard JK, Stephens P, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Saghai Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer—length polymorphism in barley: Mendelian inheritance, chromosomal location and population dynamics. *Proc Natl Acad Sci (USA)* 81:8014–8018
- Song QJ, Marek LF, Shoemaker RC, Lark KG, Concibido VC, Delannay X, Specht JE, Cregan PB (2004) A new integrated genetic linkage map of the soybean. *Theor Appl Genet* 109(1):122–128
- Tomar J, Saini N, Goyal BS, Tripathi N, Shrivastava AN, Verma RK, Tiwari S (2011) Assessment of genetic diversity among Rhizoctonia root rot resistant soybean. *J Food Leg* 24:267–272
- Torres AR, Grunvald AK, Martins TB, Santos MAD, Lemos NG, Silva LAS, Hungria M (2015) Genetic structure and diversity of a soybean germplasm considering biological nitrogen fixation and protein content. *Sci Agric* 7:47–52
- Velusamy V, Lee KJ, Ha BK, Kim JB, Kim SH, Ahn JW, Kang SY, Kim DS (2013) Evaluation of genetic diversity in Korean soybean landraces by protein banding patterns using high-throughput screening. *J Crop Sci Biotechnol* 16:189–195
- Vieira ESN, Schuster I, Silva RB, Oliveira MAR (2009) Variabilidade genética em cultivares de soja determinada com marcadores microssatélites em gel de agarose. *Pesq Agropec Bras* 44:1460–1466

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.