Molecular characterization and genetic diversity analysis of soybean (*Glycine max* (L.) Merr.) germplasm accessions in India

Giriraj Kumawat · Gourav Singh · C. Gireesh · M. Shivakumar · Mamta Arya · Dinesh K. Agarwal · Syed Masroor Husain

Received: 21 August 2014 / Accepted: 7 October 2014 / Published online: 25 October 2014
© Prof. H.S. Srivastava Foundation for Science and Society 2014

Abstract Molecular characterization and genetic diversity among 82 soybean accessions was carried out by using 44 simple sequence repeat (SSR) markers. Of the 44 SSR markers used, 40 markers were found polymorphic among 82 soybean accessions. These 40 polymorphic markers produced a total of 119 alleles, of which five were unique alleles and four alleles were rare. The allele number for each SSR locus varied between two to four with an average of 2.97 alleles per marker. Polymorphic information content values of SSRs ranged from 0.101 to 0.742 with an average of 0.477. Jaccard's similarity coefficient was employed to study the molecular diversity of 82 soybean accessions. The pairwise genetic similarity among 82 soybean accessions varied from 0.28 to 0.90. The dendrogram constructed based on genetic similarities among 82 soybean accessions identified three major clusters. The majority of genotypes including four improved cultivars were grouped in a single subcluster IIIa of cluster III, indicating high genetic resemblance among soybean germplasm collection in India.

Keywords Soybean · SSR · Germplasm · Genetic diversity · Allele

Electronic supplementary material The online version of this article (doi:10.1007/s12298-014-0266-y) contains supplementary material, which is available to authorized users.

G. Kumawat (✉) · C. Gireesh · M. Shivakumar · M. Arya · D. K. Agarwal · S. M. Husain
Directorate of Soybean Research, Khandwa Road,
Indore, M. P., India
e-mail: girirajbte@gmail.com

G. Singh
Barkatullah University, Bhopal, M. P., India

Introduction

Soybean (*Glycine max* (L.) Merrill) cultivation in India dates back to the 1st century AD (Hymowitz 1990), however commercial cultivation of soybean in India started only few decades ago with unprecedented growth in the cultivated area and total production (Tiwari et al. 1999; Agarwal et al. 2013). In the present scenario of oilseed production in India, soybean is the leading oilseed crop which is grown over 10.69 million hectare with total production of 14.66 million tonnes during the year 2012 (Anonymous 2014). The productivity of soybean in India was recorded 1370 Kg/ha in 2012, which is significantly lower as compared to other major soybean producing countries.

The genetic base of soybean cultivars is considered to be extremely narrow (Hymowitz 1970). Soybean being self pollinated crop with limited out crossing is highly inbreeding crop. During the last few decades, soybean breeding in India mainly focused on hybridization program using few selected genotypes as parental lines that has led to narrow genetic base. So far 108 improved varieties have been developed and released in India for cultivation. However, the untapped valuable genetic diversity of soybean is yet to be fully utilized to enhancing the soybean production and productivity and to broaden the genetic base. Therefore, understanding the genetic diversity of soybean germplasm is essential to broaden the genetic base and to further utilize it in breeding program. Such an insight could be achieved through molecular characterization of soybean germplasm using DNA markers, which are more informative, stable and reliable, as compared to conventional methods like pedigree analysis and morphological diversity. Early studies have shown utilization of molecular markers for identification of genetically diverse genotypes to use in crosses in breeding programme (Maughan et al. 1996; Thompson and Nelson 1998).

Directorate of Soybean Research (DSR) India, designated as National Active Germplasm Site (NAGS) for Soybean,

holds 4248 germplasm accessions of soybean which comprises of indigenous collections, land races, wild species and exotic collections from USA, Taiwan, Philippines, China, Brazil, Argentina and Thailand (Prabhakar and Bhatnagar 1995; Agarwal et al. 2013). Although morphological characterization of the whole soybean germplasm collection of India has been done but molecular characterization, a useful method for understanding extent of variation and relationship among different accessions is not performed yet. Among different types of DNA markers being utilized for molecular characterization and genetic diversity analysis in plants, simple sequence repeats (SSR) markers are considered as molecular marker of choice due to their abundance, high polymorphism rate and high reproducibility. SSR markers have been widely used in the genetic diversity studies of the soybean germplasm collections worldwide and high levels of polymorphism at SSR loci have been reported for both the number of alleles per locus and the gene diversity (Maughan et al. 1995; Abe et al. 2003; Wang et al. 2006, 2010; Fu et al. 2007; Wang and Takahata 2007; Li et al. 2008; Singh et al. 2010; Tantasawat et al. 2011).

Since, the soybean germplasm collection in India has been acquired from many different countries, their genetic relatedness and gene diversity is unknown. In India, Singh et al. (2010) studied genetic diversity in a set of 44 genotypes including 29 germplasm accessions and 15 cultivars differing in photoperiod response, using SSRs and amplified fragment length polymorphism (AFLP) markers. But, as the selected germplasm accessions were trait specific and small in number, this may not provide true picture of genetic diversity in soybean germplasm collection of India. Therefore in the present study a set of 82 soybean accessions were selected randomly from the whole germplasm collection for molecular characterization and genetic diversity analysis using SSR markers. Four improved cultivars of India were also included in this set, to assess their relative genetic diversity with respect to germplasm collection of India.

Material and methods

Plant material

A total of 82 soybean accessions obtained from NAGS for Soybean, Directorate of Soybean Research (DSR), Indore, Madhya Pradesh, were used in the present study. These 82 accessions comprises of four soybean varieties namely JS335, JS97-52, JS95-60 and Type49, and 78 soybean germplasm accessions. Out of 78 germplasm accessions, 31 are indigenous collections and remaining are originally sourced from different countries viz., 19 from USA, seven from Taiwan, five from Brazil, two each from Australia, China, Sri Lanka and one each from Philippines, Italy, Russia, South Africa, Thailand,

Germany, Argentina, Fiji, Hungary, Nepal and Japan, at different times (Prabhakar and Bhatnagar 1995). The details of 82 accessions of soybean, their country of acquisition/origin and maturity duration are given in Table 1.

DNA extraction and PCR amplification

Total genomic DNA extracted from young fresh leaves following standard CTAB extraction protocol (Doyle and Doyle 1990) was used for polymerase chain reaction (PCR) amplification of SSRs. The SSR markers developed by Cregan et al. (1999) were used in the present study. A total of 44 different SSR markers from all 20 linkage groups of soybean were chosen for genotyping of 82 accessions of soybean germplasm. The details of SSR markers, their sequences and motifs are given in Supplementary Table 1.

PCR reaction was performed in 20 µl volume of PCR mixture, containing 50 ng genomic DNA, 1X *Taq* DNA polymerase buffer, 1.5 mM MgCl₂, 0.5 mM dNTPs, 0.5U DreamTaq DNA polymerase (Thermo Scientific, USA) and 10 pmol of each primer. Thermal profiling was setup with initial denaturation temperature of 95 °C for 05 min followed by the 35 cycle of denaturation (95 °C for 60 s), annealing (55 °C for 60 s) and extension (72 °C for 90 s). The amplified SSR fragments were size separated on metaphor agarose (3.5 %) gels (Lonza, Switzerland) containing ethidium bromide (0.5 µg/ml) in 1X TAE buffer and photographed on UV transilluminator (SynGene, UK).

SSR allele scoring and data analysis

The presence or absence of SSR fragment in each accession was recorded for all the polymorphic SSR markers. The SSR bands appearing without ambiguity were scored as 1 (present) and 0 (absent) for each primer. The size of the amplified product was calculated on the basis of its mobility relative to molecular mass of marker (50 bp DNA ladder, Thermo Scientific, USA). The polymorphism information content (PIC), a measure of the allelic diversity at a locus, was determined by using the formula described by Botstein et al. (1980)

$$PIC = 1 - \sum P_i^2$$

Where, P_i is the frequency of the i^{th} allele in the set of genotypes analyzed, calculated for each SSR locus. The genetic similarity among genotypes was estimated based on Jaccard's similarity coefficient. The resulting similarity matrix was further analysed using the unweighted pair-group method arithmetic average (UPGMA) clustering algorithm for construction of dendrogram; the computations were carried out using NTSYSpc version 2.2 (Rohlf 2000).

Table 1 Details of 82 soybean accessions used for molecular characterization and genetic diversity analysis

S. No.	Accession name	Country of acquisition/origin	Maturity Duration
1	EPS71	India	97
2	BR38	Brazil	86
3	Mercury	USA	80
4	PI497966	USA	84
5	EC457254	USA	84
6	JSM258	India	86
7	EC65772	Unkown	84
8	ACC09171	Srilanka	98
9	PI170890	South Africa	98
10	IC4882	India	97
11	KTSurajSingh	India	98
12	JS97-52	India	95
13	JS95-60	India	83
14	JS20-35	India	83
15	CF-492	USA	84
16	JS20-78	India	83
17	AMSMB5-18	India	91
18	PI174859	USA	98
19	AGS44	Taiwan	83
20	BR23	Brazil	94
21	EC329156	USA	94
22	EC457286	USA	94
23	SL983	India	98
24	PI535807	USA	96
25	EC287754	Philippines	91
26	Doko Rc	Brazil	98
27	EC34057	Hungary	94
28	PI219698	USA	94
29	KS4895	USA	86
30	AGS150	Taiwan	95
31	AGS314	Taiwan	82
32	JS82	India	95
33	PM13	Srilanka	94
34	MAUS54	India	95
35	JS20-74	India	84
36	KTRameshwar	India	91
37	JS20-72	India	81
38	AGS381	Taiwan	97
39	EC39751	Italy	85
40	Kohime	Taiwan	77
41	PRAB-1	India	99
42	PRAB-2	India	91
43	PI88783	China	91
44	EC37183	Russia	98
45	EC458385	Unkown	98
46	PI175192	USA	85
47	Tunia	Thailand	95

Table 1 (continued)

S. No.	Accession name	Country of acquisition/origin	Maturity Duration
48	AGS751	Taiwan	95
49	AMS39-2-1	India	90
50	DT21	Australia	97
51	PI175198	USA	83
52	TN-4-94	USA	94
53	Soja savanna	Brazil	98
54	AMS195	India	91
55	PK765	India	95
56	JS20-55	India	87
57	EC15656	Germany	98
58	EC251519	Argentina	99
59	G00070	Brazil	87
60	EC18593	Nepal	98
61	SC15	India	98
62	MACS303	India	95
63	EC114527	Fiji	97
64	MAUS55	India	97
65	RICUM	Japan	96
66	Doles	USA	86
67	GC84051-9-1	Taiwan	86
68	EC457161	USA	98
69	JS335	India	91
70	JS20-48	India	93
71	EC39536	Unkown	96
72	Type49	India	99
73	GP116	India	93
74	PI70076	China	82
75	KS4694	USA	84
76	JS20-85	India	93
77	EC333915	USA	93
78	GP465	India	93
79	EC456534	USA	84
80	CM-60	Australia	96
81	JS20-50	India	93
82	EC289099	USA	96

Results and discussion

Molecular characterization of soybean accessions

Molecular characterization of germplasm accessions reveals underlying allelic diversity and genetic base of germplasm collection. In the present study out of total 44 SSRs analyzed, 40 were found polymorphic among 82 soybean accessions. The 82 accession of soybean were profiled with 40 polymorphic SSR markers which produced 119 alleles. The details of SSR loci, their allele number, PIC values and allele size range

have been provided in Table 2. The allele number for each SSR locus varied from two to four with an average of 2.97. The fragment size of these 119 alleles was ranged from 70 to

335 bp. Figure 1 showed an example of DNA profiles at the satt184 loci with 4 distinct alleles among different soybean genotypes. Out of the 119 alleles amplified, unique alleles

Table 2 Details of 44 SSR loci used in the study showing number of alleles, PIC value and allele sizes in 82 soybean accessions

S.No.	Linkage Group	SSR Name	Number of Alleles	PIC Value	Allele Size range (bp)
1	A1	Satt155	3	0.384	175–200
2	A1	Satt200	2	0.501	190–200
3	A1	Satt717	2	0.418	240–250
4	A2	Sat_406	3	0.587	70–200
5	A2	Sat_409	3	0.667	150–200
6	B1	BE806308	3	0.405	150–200
7	B1	Satt484	2	0.101	290–300
8	B2	Satt126	2	0.501	70–100
9	B2	Satt687	3	0.642	200–220
10	C1	Satt164	2	0.400	230–250
11	C1	Satt396	2	0.474	230–240
12	C2	Sat_252	2	0.445	150–200
13	C2	Satt557	3	0.518	170–200
14	D1a	Satt184	4	0.681	130–160
15	D1a	Satt129	4	0.742	120–155
16	D1a	Satt077	monomorphic	–	–
17	D1b	Sat_227	3	0.4809	240–255
18	D1b	Satt459	2	0.168	190–200
19	D2	Satt328	monomorphic	–	–
20	D2	Satt310	2	0.495	200–230
21	E	Satt411	3	0.495	100–150
22	E	Satt230	2	0.215	150–170
23	F	Sat_390	4	0.690	240–255
24	F	Satt362	3	0.541	250–265
25	G	Satt163	3	0.593	240–260
26	G	Satt517	3	0.427	200–270
27	H	Satt666	3	0.211	230–250
28	H	Sat_218	3	0.399	260–305
29	I	Satt127	monomorphic	–	–
30	I	Satt270	3	0.593	185–205
31	I	Satt571	3	0.467	110–115
32	J	Satt414	4	0.401	300–335
33	J	Sat_393	3	0.515	300–330
34	K	Sat_167	3	0.356	175–200
35	K	Satt055	4	0.310	70–115
36	K	Satt588	4	0.671	90–165
37	L	Sat_286	3	0.652	155–205
38	L	Satt076	monomorphic	–	–
39	M	Sat_316	3	0.530	240–300
40	M	Sat_276	4	0.600	230–300
41	N	Sat_195	3	0.184	140–150
42	N	Satt022	3	0.582	160–180
43	O	Sat_196	4	0.568	200–260
44	O	Sat_190	4	0.504	120–190
			Total=119	Average = 0.477	

were amplified by five different SSR loci in five genotypes (Table 3). Four SSR alleles were found rare with a frequency of less than 0.05 in the whole sample studied (Table 4). The two genotypes PI175192 and CF-492 amplified unique alleles as well as rare alleles. These two genotypes may serve as good sources for identification of new alleles of important genes. Two of the SSR markers amplifying unique alleles were known to be linked with two different phenotypic traits. The Satt055 was reported to be linked with photoin sensitivity trait (Liu et al. 2011). The Satt055 had amplified a unique allele in accessions EC333915. Another marker Satt196 known to be linked with yield QTLs (Du et al. 2009) had amplified a unique allele in accession CF-492. The genotypes identified for photoin sensitivity and yield QTL marker alleles can be used in marker assisted introgression program but further validation is required for marker traits linkage in segregating populations.

PIC values, a measure of the allelic diversity of SSRs, ranged from 0.101 in satt285 to 0.742 in satt129 with an average of 0.477. The high rate of polymorphic SSR loci (90.9 %) detected in this study was consistent with previous studies (Diwan and Cregan 1997; Narvel et al. 2000; Singh et al. 2010; Tantasawat et al. 2011). 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 (Cregan et al. 1999). However, the lower allele number and PIC values indicates low allelic diversity in present set of soybean accessions. The SSR allelic diversity detected among soybean genotypes in this study was low compared to most previous studies (Diwan and Cregan 1997; Narvel et al. 2000; Abe et al. 2003; Wang et al. 2006; Guan et al. 2010; Tantasawat et al. 2011), but comparable with Singh et al. (2010) who detected an average of 3.53 alleles per locus in 44 soybean genotypes from India differing in photo-period response, using 40 SSR markers.

Fig. 1 DNA profiles of 82 soybean accessions at the Satt184 loci. bp= base pairs, M =50 bp DNA ladder, 1–82 soybean germplasm accessions as per name given in Table 1, X = random genotype samples not included in data analysis

Table 3 Details of five unique SSR alleles identified in the set of 82 soybean accessions

S. No.	SSR name	Unique allele Size (bp)	Genotype showing unique allele
1	Satt666	230	PI175192
2	Satt055	70	EC333915
3	Sat_196	270	CF-492
4	Sat_190	150	EPS71
5	Sat_276	220	DT21

Molecular diversity analysis

Knowledge on the genetic diversity of germplasm helps in selection of parental genotypes for development of segregation population and to develop varieties. Morphological traits based genetic diversity is prone to environmental variations and availability of limited number of morphological markers has limited their use in genetic diversity studies. On the other hand, molecular markers based genetic diversity is not influenced by environmental factors therefore highly reproducible and also widely distributed throughout the genome. To efficiently broaden the genetic base of modern soybean cultivars, an insight into molecular diversity is necessary. In the present study, an attempt has been made to study the molecular diversity of 82 accessions using 40 polymorphic SSR markers. The high polymorphism observed in SSR markers among the soybean accessions in the present study demonstrated the effectiveness of SSR markers in determining genetic variation.

To assess the genetic resemblances among the genotypes, Jaccard's similarity coefficients were calculated for all 119 SSR alleles detected among 82 accessions. The pairwise genetic similarity among 82 soybean accessions varied from 0.28 to 0.90. The similarity coefficients matrix was used for UPGMA cluster analysis. The dendrogram constructed based on genetic similarities between genotypes showed that the 82 genotypes formed three major clusters (Fig. 2). The genetic structure of the 82 soybean accessions based on the dendrogram does not correlate with country of acquisition or morphological differentiation based on maturity duration. The Cluster I contains four genotypes namely Kohime from Taiwan, EC37183 from Russia, PI175192 and EC458385 from USA. The Cluster II also contains four genotypes namely JS20-78 and AMS39-2-1

Table 4 Details of four rare SSR alleles identified in the set of 82 soybean accessions

S. No.	SSR name	Rare allele Size (bp)	Genotypes showing rare allele
1	Satt414	335	EC457254, PI175192
2	Satt055	110	IC4882, Tunia
3	Sat_167	200	CF-492, PI174859, KS4895
4	Sat_195	160	CF-492, Kohime, KS4694

from India, AGS751 from Taiwan and DT21 from Australia. The Cluster III is largest consisting of 74 genotypes, which is further subdivided into two subclusters namely IIIa and IIIb. The subcluster IIIb contains six genotypes, of which three are

from India, one from Brazil, one from USA and one from China. The subcluster IIIa contains rest of 68 soybean genotypes. The closest related genotypes in the subgroup IIIa were Type49 and GP116, at similarity coefficient of 0.90 followed

Fig. 2 Dendrogram showing genetic relationships among 82 soybean accessions based on UPGMA clustering of Jaccard's similarity coefficients

by EC15656 and EC251519 at similarity coefficient of 0.89. Four improved cultivars included in this study were present in a single subcluster IIIa, indicating narrow genetic base of Indian soybean cultivars. While the cultivar JS95-60 was closely related to accession JS20-55 at similarity coefficient of 0.62, the cultivar JS97-52 was related to accessions PRAB-2 and MAUS54 at similarity coefficient of 0.76. The popular adapted cultivar JS335 was found related to PI219698 at similarity coefficient of 0.64.

The clustering of large number of soybean germplasm accessions in a single cluster indicates that soybean germplasm collection of India at NAGS has high genetic relatedness among accession, barring few accessions possessing diverse genetic background i.e. genotypes of the cluster I & II. While the narrow genetic diversity identified among soybean germplasm collection of India in present study necessitates the need of broadening genetic diversity by introducing more exotic germplasm along with utilization of wild relatives, the diverse accessions identified in this study may serve as source of new alleles in soybean breeding programme of India.

References

- Abe J, Xu DH, Suzuki Y, Kanazawa A, Shimamoto Y (2003) Soybean germplasm pools in Asia revealed by nuclear SSRs. *Theor Appl Genet* 106:445–453
- Agarwal DK, Billore SD, Sharma AN, Dupare BU, Srivastava SK (2013) Soybean: introduction, improvement, and utilization in India—problems and prospects. *Agric Res* 2:293–300
- Anonymous (2014) Annual report 2013–14, AICRP on Soybean, Directorate of Soybean Research, Indore, India.
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphism. *Am J Hum Genet* 32:3
- 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
- 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
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Du W, Wang M, Fu S, Yu D (2009) Mapping QTLs for seed yield and drought susceptibility index in soybean (*Glycine max* L.) across different environments. *J Genet Genom* 36: 721–731
- 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
- 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
- Hymowitz T (1970) On the domestication of the soybean. *Econ Bot* 24: 408–421
- Hymowitz T (1990) Soybeans: the success story. In: Janick J, Simon JE (eds) *Advances in new crops: proceedings of the first national symposium on new crops: research, development, economics*, held at Indianapolis, Ind., 23–26 October 1988. Timber Press, Portland, pp 159–163
- Li W, Han Y, Zhang D, Yang M, Teng W, Jiang Z, Qiu L, Sun G (2008) Genetic diversity in soybean genotypes from north-eastern China and identification of candidate markers associated with maturity rating. *Plant Breed* 127:494–500
- Liu et al (2011) QTL identification of flowering time at three different latitudes reveals homeologous genomic regions that control flowering in soybean. *Theor Appl Genet* 4:545–553
- Maughan PJ, Saghai Maroof MA, Buss GR (1995) Microsatellite and amplified sequence length polymorphism in cultivated and wild soybean. *Genome* 38:715–725
- Maughan PJ, Saghai Maroof MA, Buss GR, Huestis GM (1996) Amplified fragment length polymorphism (AFLP) in soybean: species diversity, inheritance and near isogenic line analysis. *Theor Appl Genet* 93:392–401
- 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
- Prabhakar, Bhatnagar PS (1995) Evaluation of soybean genetic resources. NRCS, Indore pp 1–96
- Rohlf FJ (2000). NTSYS-pc: Numerical taxonomy and multivariate analysis system, version 2.2. Exeter Software, Setauket, New York, USA.
- Singh RK, Bhatia VS, Bhat KV, Mohapatra T, Singh NK, Bansal KC, Koundal KR (2010) SSR and AFLP based genetic diversity of soybean germplasm differing in photoperiod sensitivity. *Genet and Mol Bio* 33:319–324
- 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. *Aust. J. Crop Sci* 5:283–290
- Thompson JA, Nelson RL (1998) Utilization of diverse germplasm for soybean yield improvement. *Crop Sci* 38:1362–1368
- Tiwari SP, Joshi OP, Sharma AN (1999) The advent and renaissance of soybean: a landmark in Indian agriculture. NRCS, Indore, pp 1–58
- Wang KJ, Takahata Y (2007) A preliminary comparative evaluation of genetic diversity between Chinese and Japanese wild soybean (*Glycine soja*) germplasm pools using SSR markers. *Genet Resour Crop Evol* 54:157–165
- Wang L, Guan R, Zhangxiong L, Chang R, Qiu L (2006) Genetic diversity of Chinese cultivated soybean revealed by SSR markers. *Crop Sci* 46:1032–1038
- Wang M, Li R, Yang W, Du W (2010) Assessing the genetic diversity of cultivars and wild soybeans using SSR markers. *African J of Biotechnol* 9:4857–4866