J. Crop Sci. Biotech. 2017 (September) 20 (3) : 221 ~ 231
DOI No. 10.1007/s12892-017-0023-0

RESEARCH ARTICLE

Genetic Diversity and Population Structure of Indian Soybean [*Glycine Max* (L.) Merr.] Revealed by Simple Sequence Repeat Markers

SK Gupta*, JG Manjaya

Nuclear Agriculture and Biotechnology Division, Bhabha Atomic Research Centre, Trombay, Mumbai-400 085, India

Received: April 30, 2017 / Revised: July 24, 2017 / Accepted: August 10, 2017

Abstract

Genetic diversity in 90 Indian soybean cultivars was assessed using 45 SSR markers distributed on 20 soybean chromosomes. Forty-five SSR markers generated 232 alleles with an average of five alleles/locus. The observed frequencies of the 232 alleles ranged from 0.01 to 0.94 with an average of 0.19. The polymorphic information content (PIC) value of the SSR markers varied from 0.10 to 0.83 with an average of 0.61 and about 71% markers have a PIC value of >0.5. In this study, 54 rare alleles including 19 genotype specific alleles were also identified. The observed hetrozygosity for SSR markers ranged from 0 to 0.11 with a mean of 0.10. Cluster analysis grouped the 90 soybean cultivars into three major clusters and principal coordinates analysis (PCoA) results were similar to those of the cluster analysis. A combination of eight SSR markers successfully differentiated all 90 soybean cultivars. The population structure analysis distributed the 90 soybean genotypes into two populations with mean alpha (α) value of 0.1873. In AMOVA analysis, proportion of variation within population was high (88%), whereas only 12% occurred among populations. In cluster and structure analyses, most of the genotypes with similar pedigree were grouped together. Soybean cultivars DS228, MACS-13, LSb-1, Hardee, Improved Pelican, and Pusa-24 were the six most genetically distinct cultivars identified. The study reported a moderate genetic diversity in Indian soybean cultivars and findings would be useful to the soybean breeders in selecting genetically distinct parents for a soybean improvement program.

Key words : Genetic diversity, soybean, SSR marker, structure analysis

Introduction

Soybean [*Glycine max* (L.) Merr.] is the world's largest oilseed crop and accounts for about 56% of edible oilseed production. Soybean is used both as a food and fodder crop because of its high protein content (40%) on a dry matter basis. Soybean is cultivated widely in the USA, Brazil, Argentina, China, and India. In India, soybean had been traditionally grown for ages in the northern parts of India (Hymowitz 1969); however, it did not establish as a main crop in the plains for many years owing to improper cultivation practices, unavailability of good varieties, lack of processing and marketing facilities, and inadequate knowledge about its utilization (Singh 2006). The commercial cultivation of soybean in India started after the launch of an AIRCP on Soybean in 1967 by the ICAR (Tiwari 2014). The soybean varieties in India have been developed from both introduced varieties and native local landraces, and about 108 improved varieties has been released for commercial cultivation (Kumawat et al. 2015). At present, soybean is a leading oilseed crop in India with 14.67 million tons production from an approximate area of 10.84 million hectares. The average yield of soybean in India is about 1353 kg/hectare which is significantly lower in comparison to the world average (Anonymous 2014). Narrow genetic diversity in Indian soybean is considered as one of the main reasons for low average yields of the Indian soybean cultivars.

Inbreeding and evolutionary events such as domestication can results in reduction of genetic diversity and modify the allele frequencies in the populations (Halliburton 2004). The study of genetic variation is a prerequisite and vital step in

SK Gupta (✉)
Email: gupta_sk@hotmail.com

the improvement of any crop plants and is important in identifying suitable parents to develop segregating populations with maximum genetic diversity (Barrett and Kidwell 1998; Thompson et al. 1998). The extent of genetic diversity in germplasm is usually determined by studying various morphological characters (Bar-Hen et al. 1995; Manjaya and Bapat 2008). However, most of these characters are quantitative in nature and are highly influenced by environmental conditions and therefore need multilocation data for concluding the results. On the contrary, DNA or molecular markers are plenty in number and remain unaffected by environmental conditions and growth stages of the plant. Different types of molecular markers like RFLP (Botstein et al. 1980), RAPD (Williams et al. 1990), ISSR (Zietkiewicz et al. 1994), AFLP (Vos et al. 1995), and microsatellite or SSR (Tautz and Renz 1984) have been utilized for genetic analysis in crop plants. Among them, SSR which are well distributed in the eukaryote genomic regions are much preferred because of their locus specificity, reproducibility, high polymorphism rate, and codominant nature. SSRs have been used for studying genetic diversity in many legumes including soybean (Brown-Guedira et al. 2000; Chauhan et al. 2015), pigeonpea (Dutta et al. 2011), cowpea (Asare et al. 2010; and Gopalakrishna 2010), mungbean (Gupta et al. 2013), common bean (Burle et al. 2010), and blackgram (Gupta and Gopalakrishna 2009).

Occurrence of low genetic diversity in cultivated soybean germplasm is a concern worldwide (Hyten et al. 2006). Like many other crops, breeders are frequently using only few selected genotypes for improving soybean because of their good combining ability, dominating yield, and agronomic performance that can lead to narrow genetic base. The efficient use of unexploited genetic variability available in the germplasm can help overcome the problem of narrow genetic base and simultaneously improve the average soybean productivity. The objective of this study was to assess the genetic diversity and population structure in cultivated Indian soybean genotypes using SSR markers.

Materials and Methods

Plant materials

Ninety soybean cultivars collected from different locations in India were used and are listed in Table 1 along with their pedigree. Genomic DNA of each cultivar was isolated from fresh leaves using DNeasy® Plant Mini Kit (Qiagen, Germany). The DNA was quantified using NanoDrop ^{TM} 1000.

SSR analysis

Forty-six SSR primer pairs used in the study were randomly selected from 20 linkage groups of soybean (Song et al. 2004). The details of the SSR primer sequences are given in Supplementary Table S1. PCRs were performed using an Eppendorf Mastercycler in a 20 µl reactions volume having 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 40 ng of template DNA, 20

Table 1. Details of the soybean cultivars used in the study.

S. no.	Varieties	Pedigree	Cultivation zone/state*
1	ADT-1 (UGM-33)	Selection from "Hill" variety	TN
2	Alankar	D 63-6094 x D 61-4249	NPZ
3	Birsa Soya-1	Mutant of Sepaya black	Jharkhand
4	Bragg	Jackson×D 49-2491	All India
5	Co-1	Selection from EC 39821	TN
6	Co-Soya-2	UGM 21 x JS335	TN
7	Co-Soya-3	UGM 69 x JS335	TN
8	JS72-280 (Durga)	EC 14437×Bragg	MP
9	DS-228	JS335 X DS-181 (PI 4623133)	MS
10	DS98-14	Bragg x DS93-MN-39	NPZ
11	DS97-12	Mutant of DS74	NPZ
12	Gaurav	EC1143 x Bragg	MP
13	Guj Soya-1 (J-231)	Selection form Punjab-1 variety	Gujarat
14	Guj Soya-2 (J-202)	Selection from Geduld variety	Gujarat
15	Hardee	D49-772 x Improved Pelican	SZ
16	Hara Soya (Himso 1563)	Himso-1520 × Bragg	HP, Uttranchal
17	Indira Soy 9	Selection from JS 80-21	NEZ, MP
18	Improved Pelican	Tanloxi ×P.I.60406	SZ
19	JS-2	Selection from Tehri-Garhwal material.	MP
20	JS71-05	Selection from Lee type exotic material	MP

Table 1. Details of the soybean cultivars used in the study (continued).

S. no.	Varieties	Pedigree	Cultivation zone/state*
21	JS75-46	Improved Pelican× Semmes	MP
22	JS76-205	Bragg × Kalitur	MP
23	JS79-81	Bragg × Harasoy	MP
24	JS80-21	JS75-1×PK 73-49	CZ, NEZ
25	JS90-41	PS73-7 × Hark	MP
26	JS97-52	PK327 x L129	CZ, NEZ
27	JS335	JS-78-77 ×JS 75-1	CZ
28	DSb-21	JS335 x EC241778	Karnataka, SZ
29	Kalitur	Indigenous variety	MP
30	KB-79 (Sneh)	Hardee ×Monetta	Karnataka
31	KHSb-2	Mamloxi x EC39821	Karnataka
32	Lee	S-100 × CNS	NHZ
33	LSb-1	Selection form MACS-330	Andhra Pradesh
34	MACS-13	Hampton× EC7034	Maharashtra
35	MACS-57	JS-2× Improved Pelican	Maharashtra
36	MACS-58	JS-2× Improved Pelican	CZ
37	MACS-124	JS-2× Improved Pelican	SZ
38	MACS-450	Bragg x MACS-111	SZ
39	MAUS-1	Mutant from DS 87-14	MS
40	MAUS-2 (Pooja)	Selection from SH 84-14	SZ
41	MAUS-32 (Prasad)	Selection from JS-80-21	MS
42	MAUS-47	PS 73-7 × Hark	CZ
43	MAUS-61 (Pratikar)	JS71-1 x PK-73-94	SZ
44	MAUS-61-2 (Pratishta)	JS80-21 x KB-60	CZ
45	MAUS-71 (Samrudhi)	JS71-5 x JS87-38	NEZ
46	MAUS-81 (Shakti)	KB-74 x JS335	CZ
47	Monetta	An exotic variety EC 2587	MS
48	NRC-2 (Ahilya-1)	Induced mutant of Bragg	MP
49	NRC-7 (Ahilya-3)	Selection from S69-96	MP, Rajasthan,MS,UP
50	NRC-12 (Ahilya-2)	Induced mutant of Bragg	MP
51	NRC-37 (Ahilya-4)	Punjab-1 × Gaurav	CZ
52	Palam Soya	JS72-45-1 × Punjab-1	HP, Uttranchal
53	PK-262	UPSM97 x Hardee	NPZ
54	PK-308	T-31× Hardee	NPZ
55	MACS-1188	JS(SH)93-01 x MACS450	SZ
56	PK-416	UPSM534 x S38	NPZ
57	PK-471	Hardee x Punjab-1	SZ
58	PK-472	Hardee× Punjab-1	CZ
59	PK-564	(UPSM534 x Ankur) x Bragg	NPZ
60	PS-1024	PK308 x PK317	NPZ
61	PS-1029	PK-262 × PK-317	SZ
62	PS-1042	Bragg x PK 416	NPZ
63	PS-1092	PK327 x PK416	UP, Uttranchal
64	PS-1225	Pk515 x PK327	UP, Uttranchal
65	PS-1241	PK1039 x PK327	UP, Uttranchal
66	PS-1347	PS1024 x PK472	NPZ

Table 1. Details of the soybean cultivars used in the study (continued).

S. no.	Varieties	Pedigree	Cultivation zone/state*
67	Punjab-1	Selection from Nanking Variety	CPZ, NHZ
68	Pusa-16	CNS× Lee	NPZ, NHZ
69	Pusa-20	Bragg x Lee	NHZ
70	Pusa-22	Punjab × Clark 63	NHZ
71	Pusa-24	Shelby× Bragg	NPZ, NHZ
72	Pusa-37	Bragg x Java-16	NPZ, NEZ, CZ
73	RKS-18	MACS-450 x Monetta	SZ, NEZ
74	Shilajeet	Selection from EC9309	NPZ, NHZ
75	Shivalik	Selection from segregating PK73-55	HP
76	SL-295	PK416 x PS564	Punjab
77	SL-525	PK416 x PK1023	NPZ
78	SL-688	PK416 x SL317	NPZ
79	TAMS-38	Monetta x PK472	MS
80	RKS-24C	PK472 x PS1024	Rajasthan
81	Type-49	Selection from indigenous Material	CZ
82	VL Soya-1	Mutant of Bragg	Uttranchal
83	VL Soya-2	Selection from VHC856007	NHZ
84	VL Soya-21	Selection from VHC 3055	Uttranchal
85	VL Soya-47	Selection from KHSF-3-1-1	NHZ
86	VL Soya-59	(Pb-1 x VLS-2) x EC361336	NHZ
87	VL Soya-63	VLS-2 x (Bragg x VHC3022)	NHZ
88	DSb-1	Selection from EC172576	Karnataka
89	DSb-12	JS335 x PS73-7	SZ
90	Samrat	local selection	CZ

*NPZ = North plain zone; SZ = South zone; CZ = Central zone; CPZ = Central plain zone; NEZ = North east zone; NHZ = North hill zone; MS = Maharashtra; MP = Madhya Pradesh; TN = Tamil Nadu; UP = Uttar Pradesh; HP = Himachal Pradesh

ng of each primer pair, and 0.5 unit of *Taq* DNA polymerase (Jonaki, Hyderabad, India). Amplification conditions were: one cycle of 95 ^{0}C for 4min, followed by 40 cycles of 94 ^{0}C for 30 s, 55-60 ^{0}C for 30 s, 72 ^{0}C for 30 s, and one cycle of 72 ^{0}C for 10 min. Amplification products were resolved on 4% MetaPhor® agarose gels (Cambrex, USA). Gels were stained in ethidium bromide solution (0.5 ug/ml) and photographed using a digital gel documentation system. For some PCR amplification products which required higher resolution, products were separated on 6% denaturing polyacrylamide gels. For this, 5 µl of each PCR product was electrophoresed on denaturing gel using a Sequi-Gen GT Sequencing system (Bio-Rad, USA). Gels were visualized using a modified silver staining procedure (Gupta and Gopalakrishna 2010).

Statistical analyses

Scoring of marker alleles was done as presence (1) or absence (0) at each SSR locus. The allelic and genotype frequencies, expected hetrozygosity and observed hetrozygosity for each locus were evaluated using software PowerMarker version 3.25 (Liu and Muse 2005). PIC for each markers' locus was calculated by using the formula of Powell et al. (1996): PIC = $1 - \Sigma(P_{ij})^2$, where P_{ij} is the frequency of the j^{th} alleles for i^{th} loci. Hierarchical cluster analysis was performed using software DARWin ver 6.0 (Perrier et al. 2003). For calculating genetic relationship among individuals, allelic data was used to generate pairwise dissimilarities using simple matching coefficients. The dissimilarity coefficients were subjected to neighbor joining method for estimating genetic relationship among the cultivars. Bootstrap analysis with 5000 permutations was performed to provide support for clustering. Principal coordinates analysis (PCoA) was performed to highlight the resolving power of ordination using DARWin version 6.0 software.

The model-based method in STRUCTURE software version 2.3.4 (Pritchard et al. 2000) was used to study the population structure. Admixture model and correlated allele frequency model were applied to analyze the dataset with a burn-in period of 50,000 generations followed by 50,000 replications. The assumed number of populations (K) was set from 1 to 10 and for each K value, 10 runs of STRUCTURE were performed. The plateau of ΔK was obtained by plotting LnPD values derived from each K and the final population was determined using STRUCTURE HARVESTER software (Earl and vonHoldt 2012). Analysis of molecular variance AMOVA) was conducted using software GenAlEx 6.5

Table 2. Genetic diversity parameters estimated for SSR markers in 90 soybean cultivars.

S. no.	SSR locus	Total no. of alleles	Rare alleles	Major allele frequency	Heterozygosity	PIC value
1	Satt684	5	2	0.66	0.02	0.47
2	Sat_356	5	1	0.33	0.01	0.74
3	Sat_374	8	2	0.39	0.00	0.77
4	Satt424	5	0	0.53	0.01	0.65
5	Satt409	6	2	0.44	0.00	0.70
6	Sat_272	5	1	0.43	0.03	0.63
7	Satt509	3	0	0.65	0.01	0.50
8	Sat_177	9	3	0.37	0.00	0.80
9	Satt304	6	3	0.49	0.00	0.67
10	Satt687	4	1	0.77	0.00	0.39
11	Satt565	4	1	0.69	0.02	0.46
12	Satt524	2	0	0.56	0.00	0.49
13	Satt681	9	4	0.50	0.00	0.70
14	Satt357	5	1	0.61	0.00	0.56
15	Sat_413	4	0	0.36	0.00	0.73
16	Satt184	5	1	0.46	0.00	0.66
17	Satt216	5	1	0.44	0.11	0.64
18	Satt428	4	1	0.81	0.00	0.32
19	Sat_415	5	1	0.29	0.00	0.76
20	Sat_296	7	3	0.39	0.07	0.72
21	Satt014	2	0	0.83	0.01	0.27
22	Sat_022	5	1	0.48	0.00	0.67
23	Satt213	1	0	-	-	-
24	Satt231	5	0	0.56	0.00	0.63
25	Satt325	8	2	0.44	0.00	0.75
26	Sat_234	4	0	0.38	0.00	0.73
27	Satt610	4	2	0.81	0.00	0.31
28	Satt138	6	0	0.29	0.00	0.78
29	Satt191	5	0	0.28	0.00	0.78
30	Satt314	3	1	0.66	0.00	0.46
31	Sat_180	6	0	0.42	0.00	0.73
32	Satt419	5	1	0.68	0.01	0.49
33	Sct_189	4	1	0.44	0.07	0.59
34	Sat_361	7	2	0.44	0.00	0.73
35	Sat_144	3	0	0.78	0.02	0.35
36	Satt326	3	1	0.77	0.00	0.37
37	Satt588	10	4	0.35	0.01	0.81
38	Sat_340	3	1	0.52	0.00	0.53
39	Satt373	8	1	0.26	0.00	0.83
40	Satt150	7	4	0.44	0.00	0.67
41	Satt308	6	1	0.26	0.00	0.80
42	Satt152	6	3	0.53	0.00	0.63
43	Sat_275	5	1	0.43	0.00	0.71
44	Sat_306	4	0	0.40	0.00	0.71
45	Satt492	2	0	0.94	0.00	0.10
46	Sat_282	5	0	0.33	0.00	0.76

(Peakall and Smouse 2012) to calculate variance components and their statistical significance levels for variation among and within the populations.

Results

All 46 SSR markers amplified successfully across the soybean genotypes and 45 of them were found polymorphic. The SSR marker Satt213 amplified only one allele and was found monomorphic. Forty-five polymorphic SSR loci generated 232 alleles and number of alleles varied from 2 to 10 with an average of 5 alleles/locus (Table 2). The amplification pattern of three SSR markers on soybean cultivars is shown in Fig. 1. PIC value of SSRs ranged from 0.10 (Satt492) to 0.83 (Satt373) with an average of 0.61. The observed hetrozygosity for the SSR markers varied from 0 to 0.11 (Table 2). The major allele frequency for the SSR markers ranged from 0.26 (Satt308 and Satt373) to 0.94 (Satt492) with an average of 0.51. Nineteen markers with major allele frequency of more than 0.50 and four markers with major allele frequency of 0.80 or larger were identified in the study. In addition, 54 rare alleles (allele having frequency of less than 5%) from 31 SSR loci were identified. The maximum numbers of rare alleles were observed for SSR loci Satt150,

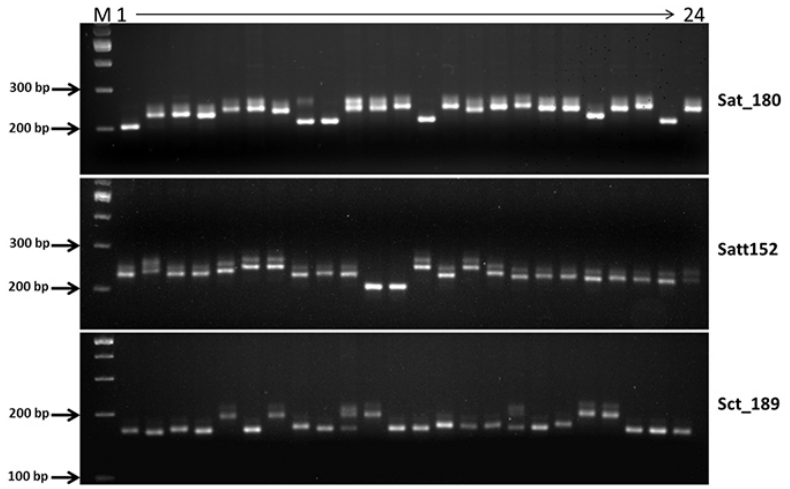

Fig. 1. PCR amplification profile generated with SSR markers (Sat_180, Satt152, and Sct_189) in 24 soybean cultivars.

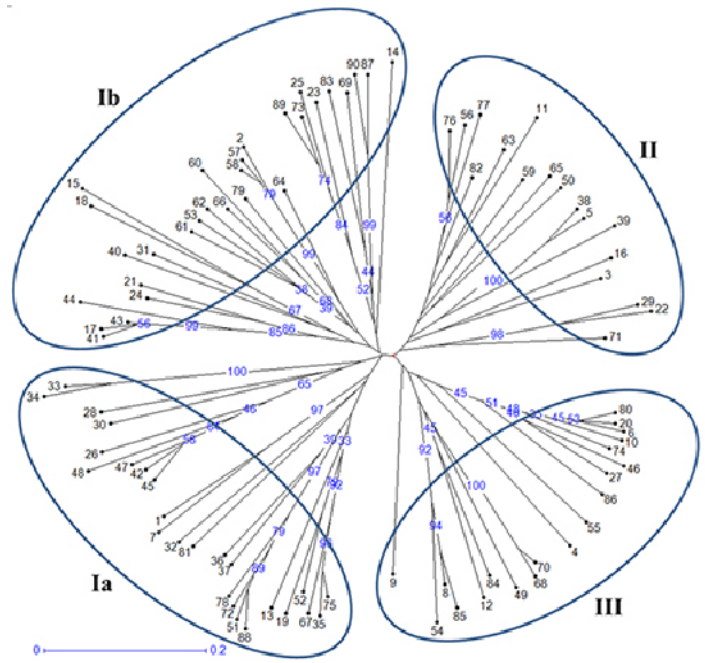

Fig. 2. Neighbo Cluster analysis grouped the 90
r-joining tree with bootstrap support values (>30%, based on 2,000 bootstraps) showing genetic relatedness among 90 soybean cultivars based on SSR markers. Numbers represent soybean cultivars as listed in Table 1.

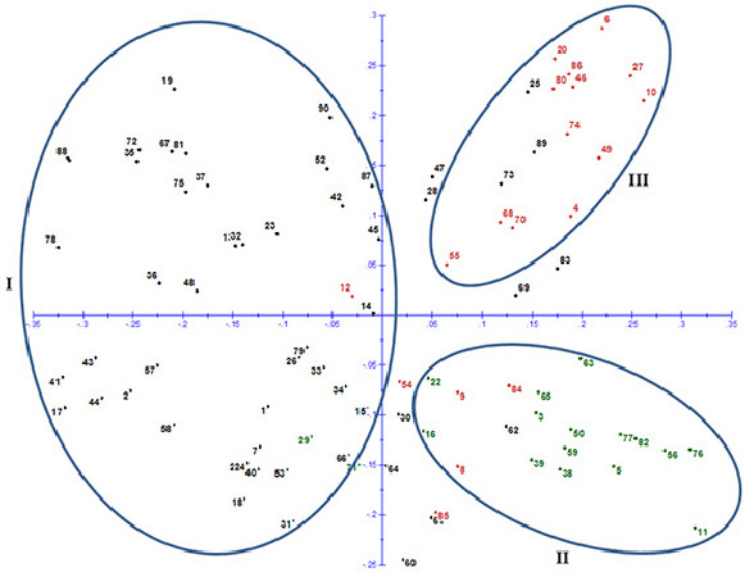

Fig. 3. Principal Coordinates Analysis (PCoA) showing relationship among soybean cultivars based on SSR markers. Numbers represent soybean cultivars as listed in Table 1 (Grouping based on hierarchical clustering approach).

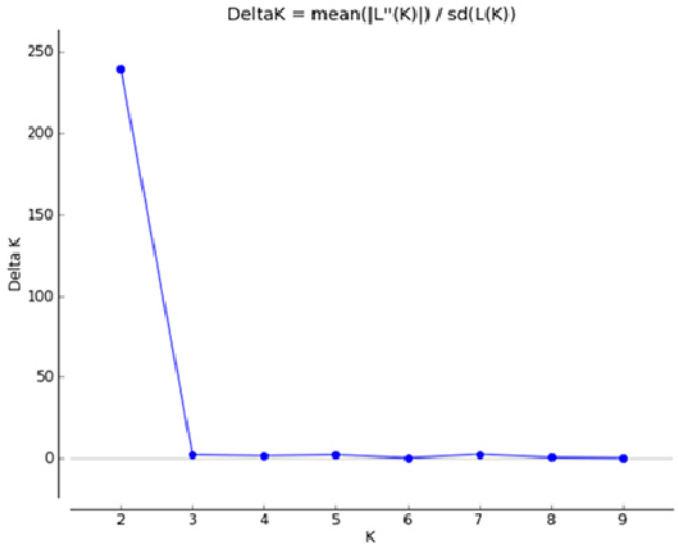

Fig. 4. Estimation of population using LnP(D) derived ΔK for K from 1 to 10.

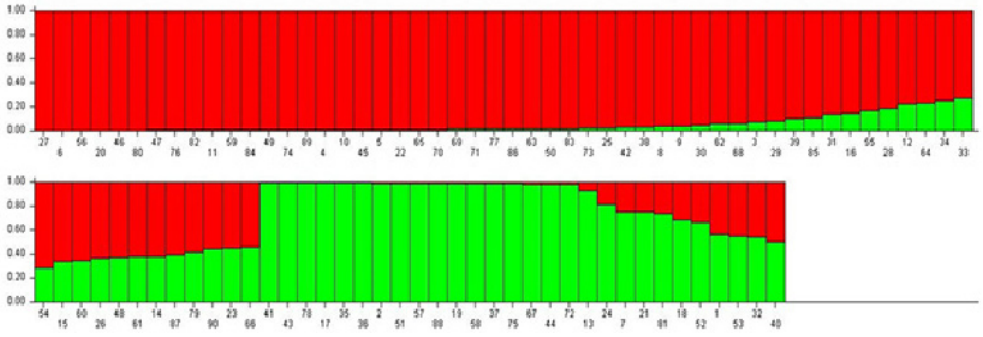

Fig. 5. Inferred population structure of 90 soybean cultivars. Each cultivar is represented by a single vertical line and numbers represent soybean cultivars as listed in Table 1. Each colour represents one cluster, and the length of the colored segment shows the cultivar's estimated proportion of membership in that cluster.

Table 3. Analysis of molecular variance (AMOVA) for 90 individuals in two populations of soybean.

Source of variation	Degree of freedom	SSD	MSD	Estimated variance	Percentage of variance
Among populations	1	329.10	329.10	7.30	12% *
Within populations	88	4652.52	52.87	52.87	88% *

*: P<0.001; SSD: sum of squared deviation; MSD: mean squared deviation

Satt588, and Satt681 with 4 rare alleles each. A combination of five SSR markers (Satt373, Sat_177, Satt325, Satt588 and Sat_413) were able to differentiate 78 cultivars out of 90 cultivars studied, while a combination of eight SSR markers (Satt373, Sat_177, Satt325, Satt588, Sat_413, Sat_234, Satt152 and Sat_361) distinguished all 90 cultivars.

The pairwise genetic dissimilarity between individuals ranged from 0.09 to 0.84. Based on cluster analysis, 90 soybean cultivars were distributed in three major clusters (Fig. 2) at a genetic distance of 0.70. Cluster I was the largest cluster and contained 54 cultivars. Cluster I was further subdivided into two subclusters, cluster Ia and cluster Ib, with 25 and 29 cultivars, respectively. Cluster II had 17 cultivars and cluster III contained 19 cultivars. In the clustering, most of the genotypes showed grouping based on their pedigree. Results of the PCoA were similar to those of the cluster analysis (Fig. 3).

The structure analysis divided the 90 soybean cultivars into two populations (Figs. 4 and 5). Population 1 consisted of 63 cultivars and population 2 had 27 cultivars. Based on membership fraction, cultivars with probability score more than 0.90 were considered as pure and less than 0.90 as admixture. The population 1 had 41 pure (65%) and 22 (35%) admixed cultivars and population 2 had 18 (67%) pure and 9 (33%) admixed individuals. Overall proportion of membership of the sample in the population 1 was 0.648 and in population 2 was 0.352. The allele frequency divergence among two populations was 0.1345 and mean value of alpha (α) was 0.1873. The mean Fst value of population 1 was 0.1305 and population 2 was 0.2652. AMOVA analysis was conducted based on populations obtained by model-based approach. In AMOVA, proportion of variation attributable to within population differences was high (88%), whereas only 12% occurred among populations (Table 3).

Discussion

The reduction in genetic diversity hinders the crop improvement as no genetic gain is achieved by using genetically similar parents in hybridization. Therefore, to observe the variation in Indian soybean cultivars, genetic diversity in 90 soybean cultivars was studied using 45 polymorphic SSR markers. These 45 SSR markers were distributed over all 20 soybean linkage groups to get the representation from all soybean chromosomes. Mapped markers have the advantage that it enables genome-wide coverage and reduces over-representation of certain chromosomal regions, thus allowing more precise estimation of genetic diversity. Many other studies had reported importance of mapped markers for analysis of genetic variation (Gupta and Gopalakrishna 2009; Karp et al. 1997; Singh et al. 2004).

In this study, 45 SSR markers generated 232 alleles with an average of five alleles/locus. The polymorphism level observed among Indian soybean cultivars was similar to those reported for Canadian (6.3 allele/locus; Fu et al. 2007) and Thailand (4.8 allele/locus; Tantasawat et al. 2011) soybean genotypes, but was less compared to Chinese (12 alleles/locus; Wang et al. 2006), and American (10 alleles/locus; Diwan and Cregan 1997) soybean genotypes. The PIC values, which represent allelic diversity and frequency, had a mean of 0.61. A total of 32 SSR markers (71%) with PIC of >0.5 were observed (Table 2), suggesting high resolving power of SSR markers used in the study. The high PIC values observed were in complete agreement with previous studies (Chauhan et al. 2015; Diwan and Cregan 1997; Kumawat et al. 2015). Four SSR markers namely, Sat_177 on LG B2, Satt588 on LG K, Satt373 on LG L and Satt308 on LG M has PIC value of $\geq$ 0.80 and were most informative SSR markers. In the study, heterozygotes were also observed at 31 SSR loci. However, observed hetrozygosity was much less (average 0.10) and this was expected because soybean is a self-pollinating crop. The numbers of repeats present in SSRs are generally directly correlated to polymorphism level of the SSR markers. However, in this study, no correlation between longer SSR repeats and higher number of alleles was observed. For example, SSR marker Satt325 with $(ATT)_{19}$ repeats produced eight alleles and the marker Satt509 with $(ATT)_{30}$ repeats produced only three alleles (Table 2). In the study, 54 rare alleles including 19 genotype specific or unique alleles were also observed and rare alleles were usually present at SSR loci having more number of alleles. Similar observations have been made in other studies (Jain et al. 2004; Li et al. 2008). In this study, all 90 soybean cultivars could be unambiguously distinguished from each other by using a combination of eight SSR markers, suggesting that SSR markers used were highly polymorphic and informative. These SSR markers along with unique and rare alleles will serve as diagnostic markers for unambiguous soybean varietal identification. The discriminatory power of SSR markers for varietal differentiations has been observed in earlier studies also (Li et al. 2001; Russell et al. 1997).

Based on cluster analysis, 90 soybean genotypes were distributed into three major clusters and moderate genetic diversity among the cultivars was observed with dissimilarity coefficient varying from 0.07 to 0.84 (Fig. 2). In clustering, cultivars showed grouping based on their pedigree and no correlation between the geographical location and grouping was observed. For example, cultivars MACS-57, MACS-58, and MACS-124 sharing common parentage were grouped

together in cluster Ia. Similarly, cultivars SL-295, SL-525, and PS-1092 having PK-416 as a common parent in their pedigree, were grouped in cluster II along with PK-416 (Fig. 2). In cluster Ia, cultivar MACS-13 showed maximum diversity, and cultivars Shivalik and MACS-57 were genetically most similar. Similarly in Cluster Ib, cultivars PK471 and PK472 with same pedigree showed most similarity, and cultivars Hardee and Improved Pelican were most diverse. In cluster II, Pusa-24 was most genetically diverse cultivar, and MACS-450 and Co-1 were most similar soybean cultivars. In Cluster III, cultivars RKS-24C and JS71-05 showed the highest, whereas cultivar GS-228 showed least genetic similarity. Results of the PCoA (Fig. 3) were similar to those of the cluster analysis and first two axis contributed for 18% of the total variation. The model based structure analysis divided the 90 soybean cultivars into two populations (Fig. 4). All soybean cultivars included in Clusters II and III in dendrogram belonged to population II and cultivars belonging to Cluster I were distributed over both populations I and = II. The small alpha values ($\alpha = 0.1873$) in structure analysis indicated that most of the cultivars are pure and only few admixtures are present in the populations. More variation within population (88%) and less (12%) among populations was observed in AMOVA analysis. This was expected because soybean is a self-pollinated crop. Grouping of most of the Indian soybean cultivars in three major clusters and only two populations showed close genetic relatedness of the accessions. This genetic relatedness may be because of repeated use of few parental combinations in Indian soybean improvement programs. For example, soybean cultivar "Bragg" has been used as a parent in developing 16 of 90 soybean cultivars used in the study (Table 1). In earlier studies also, grouping of Indian soybean genotypes into two (Bisen et al. 2015; Chauhan et al. 2015) and three clusters (Kumawat et al. 2015) has been reported, indicating the close relatedness among cultivars. Bharadwaj et al. (2002) had also reported narrow genetic base in Indian soybean varieties and attributed it to repetitive use of same parents for the development of genetic stock or breeding material.

This study divulged the presence of moderate genetic diversity among Indian soybean cultivars. However, there is a need to deploy new and diverse parents in crossing program to maintain and further increase the genetic base of Indian soybean. Soybean cultivars DS228, MACS-13, LSb-1, Hardee, Improved Pelican, and Pusa-24, which were found most genetically distinct, could be utilized in soybean varietal improvement programs. The highly polymorphic and informative SSR markers revealed in the study would serve as important genomic resources for soybean germplasm characterization.

References

Anonymous. 2014. Annual report 2013–14, AICRP on Soybean, Directorate of Soybean Research, Indore, India

Asare AT, Gowda BS, Galyuon IKA, Aboagye LL, Takrama JF, Timko MP. 2010. Assessment of the genetic diversity in cowpea (Vigna unguiculata L. Walp.) germplasm from Ghana using simple sequence repeat markers. Plant Genet. Resour. 8: 142-150

Bar-Hen A, Charcosset A, Bourgoin M, Cuiard J. 1995. Relationships between genetic markers and morphological traits in a maize inbred lines collection. Euphytica 84: 145-154

Barrett BA, Kidwell KK. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci. 38: 1261-1271

Bharadwaj CH, Satyavathi CT, Tiwari SP, Karmakar PG. 2002. Genetic base of soybean (*Glycine max*) varieties released in India as revealed by coefficient of parentage. Indian J. Agri. Sci. 72(8): 467-469

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

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331

Brown-Guedira GL, Thompson JA, Nelson RL, Warburton ML. 2000. Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. Crop Sci. 40: 815-823

Burle ML, Fonseca JR, Kami JA, Gepts P. 2010. Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. Theor. Appl. Genet. 121: 801-813

Chauhan DK, Bhat JA, Thakur AK, Kumari S, Hussain Z, Satyawathi CT. 2015. Molecular characterization and genetic diversity assessment in soybean [*Glycine max* (L.) Merr.] varieties using SSR markers. Ind. J. Biotechnol. 14: 504-510.

Diwan N, Cregan PB. 1997. Automated sizing of Xuorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95: 723-733

Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V, Gaikwad K, Sharma TR, Raje RS, Bandhopadhya TK, Datta S, et al. 2011. Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [Cajanus cajan (L.) Millspaugh]. BMC Plant Biol. 11: 17

Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4(2): 359-361

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

Gupta SK, Bansal R, Vaidya UJ, Gopalakrishna T. 2013. Assessment of genetic diversity at molecular level in mungbean (*Vigna radiata* (L.) Wilczek). J. Food Legumes 26(3 & 4): 19-24

Gupta SK, Gopalakrishna T. 2009. Genetic diversity analysis in blackgram (*Vigna mungo* (L.) Hepper) using AFLP and transferable microsatellite markers from azuki bean (*Vigna angularis* (Willd.) Ohwi & Ohashi). Genome 52: 120-128

Gupta SK, Gopalakrishna T. 2010. Development of unigene-derived SSR markers in cowpea (*Vigna unguiculata*) and their transferability to other Vigna species. Genome 53: 508-523

Halliburton R. 2004. Introduction to population genetics. Pearson/ Prentice Hall, Upper Saddle River, NJ

Hymowitz T. 1969. The soybeans of the Kumaon Hills of India. Econ. Bot. 23(1): 50-54

Hyten DL, Song Q, Zhu Y, Choi IY, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. 103: 16666-16671

Jain S, Jain R, Mccouch S. 2004. Genetic analysis of Indian aromatic and qualityrice (*Oryza sativa* L.) germplasm using panels of Xuorescently-labled microsatellite markers. Theor. Appl. Genet. 109: 965-977

Karp A, Edwards KJ, Bruford M, Funk S, Vosman B. 1997. Molecular technologies for biodiversity evaluation: opportunities and challenges. Nat. Biotechnol. 15: 625-628

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(1): 101-107

Li C-D, Fatokun CA, Ubi B, Singh BB, Scoles GJ. 2001. Determining genetic similarities and relationships among cowpea breeding lines and cultivars by microsatellite markers. Crop Sci. 41: 189-197

Li Y, Guan R, Liu Z, Ma Y, Wang L, Li L, Lin F, Luan W, Chen P, Qiu L. 2008. Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. Theor. Appl. Genet. 117: 857-871

Liu K, Muse S. 2005. Powermarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128-2129

Manjaya JG, Bapat VA. 2008. Studies on genetic divergence in soybean [*Glycine max* (L.) Merrill]. J. Oilseeds Res. 25: 178-180

Peakall R, Smouse PE. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28: 2537-2539

Perrier X, Flori A, Bonnot F. 2003. Data analysis methods. In: P Hamon P, M Seguin M, PX errier X, JC Glaszmann, Ed., Genetic diversity of cultivated tropical plants. Enfield, Science Publishers. Montpellier. pp 43-76

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed. 2: 225-238

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W. 1997. Discriminating between barley genotypes using microsatellite markers. Genome 40: 442-450

Singh BB. 2006. Success of soybean in India: The Early Challenges and Pioneer Promoters. Asian Agrihist. 10: 43-53

Singh RK, Sharma RK, Singh AK, Singh VP, Singh NK, Tiwari SP, Mohapatra T. 2004. Suitability of mapped sequence tagged microsatellite site markers for establishing distinctness, uniformity and stability in aromatic rice. Euphytica 135: 135-143

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: 122-128

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(3): 283-290

Tautz D, Renz P. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genome. Nucl. Acids Res. 12: 4127-4138

Thompson JA, Nelson RL, Vodkin LO. 1998. Identification of diverse soybean germplasm using RAPD markers. Crop Sci. 38: 1348-1355

Tiwari SP. 2014. Raising the yield ceilings in soybean - An Indian Overview. Soybean Res. 12(2): 1-43

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Friters A, Pot J, Paleman J, Kuiper M, Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucl. Acids Res. 23: 4407-4414

Wang LX, Guan RX, Liu ZX, Chang RZ, Qiu LJ. 2006. Genetic diversity of Chinese cultivated soybean revealed by SSR markers. Crop Sci. 46: 1032-1038

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18: 6531-6535

Zietkiewicz E, Rafalski A, Labuda D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics 20: 176-183

Supplementary Table S1. Details of soybean SSR markers used for studying genetic diversity in Indian soybean cultivars.

S. no.	SSR locus	LG	Repeat motif	Forward primer (5'….3')	Reverse primer (5'….3')
1	Satt684	A1	(ATT)17	GGGCTTCATTTTAGATGGAGTC	TGGAGCTCATATTCGTCACAAAG
2	Sat_356	A1	(AT)26	GCGCCGGAAAAATGTGAGAAATCATAAAA	GCGTTGCATGACTATCATTCAATCAAAAAT
3	Sat_374	A1	(AT)23	GCGTTGAAACCGTTATAAACCAACTCA	GCGCTTTATTGGCAATACTTTTAACTCACAT
4	Satt424	A2	(ATT)52	CAACCTGTATTCCACAAAAAATCTCACC	GCGCCCCAATTTGACTATAAATAAAAGT
5	Satt409	A2	(ATT)27	CCTTAGACCATGAATGTCTCGAAGATA	CTTAAGGACACGTGGAAGATGACTAC
6	Sat_272	B1	(AT)19	GCGATGGCAATATGTTTTTGAGC	GCGGCCTTGTAATTTTCCTTGTTAATGTG
7	Satt509	B1	(ATT)30	GCGCTACCGTGTGGTGGTGTGCTACCT	GCGCAAGTGGCCAGCTCATCTATT
8	Sat_177	B2	(AT)13(AC)5	AGCCACGATACACAAAAACC	GGAGGGTACAAAAGTAGCAAAGATT
9	Satt304	B2	(ATT)29	GGGTAGTGACGTATTTCATGGTC	GCGTAAAAACATTCGTTGACTACATAA
10	Satt687	B2	(ATT)9	ACCGCAACTCACTCACCTT	GCGCCCAATTAACAGAAAC
11	Satt565	C1	(ATT)19	GCGCCCGGAACTTGTAATAACCTAAT	GCGCTCTCTTATGATGTTCATAATAA
12	Satt524	C1	(ATT)14	GCGAATTATCCAAAGATACACTTAGTC	GCGGGTCTTACGAACGTGTCACATTAT
13	Satt681	C2	(ATT)20	GCGGTGCACTTGTCAATCTGTT	GCGGTGAGGCATATGTCAGTC
14	Satt357	C2	(ATT)12	CCTGAGCAATTCATACTCC	TAACCGATCCGATCCTTGACA
15	Sat_413	D1a	(AT)35	GCGCTCCCTTCTTTTCCACTGAATTGA	GCGTTTTCTCTCGGTTTCTCTCTTCTTATTA
16	Satt184	D1a	(ATT)13	GCGCTATGTAGATTATCCAAATTACGC	GCCACTTACTGTTACTCAT
17	Satt216	D1b	(ATT)19	TACCCTTAATCACCGGACAA	AGGGAACTAACACATTTAATCATCA
18	Satt428	D1b	(ATT)20	GCGTTGTGCTTAACTGGTTGATTT	GCGGACCAGCTAGTTTTTAATGTG
19	Sat_415	D1b	(AT)34	GCGATTGTAAAGTAATAAGGGTATATTG	GCGTTTTGTACTCCATATACTACCTTCATC
20	Sat_296	D2	(AT)32	GCGAGACCCATTTAATTCTCAATATCAGACA	GCGCCCGTGAATGAGTCAAACAAGTA
21	Satt014	D2	(ATT)9	TCTGGTAAACATTCAACTTTTTATTT	TCCAAATATGACATCATAAACTTCTA
22	Sat_022	D2	(AT)27	GCGGCCTTTTCTGACTGTTAA	GCGCAGTGACTAAAACTTACTAT
23	Satt213	E	(ATT)15	CCGCTTATTTCTGTCATC	AGCCAAAACCCACAA
24	Satt231	E	(ATT)32	GCGTGTGCAAAATGTTCATCATCT	GGCACGAATCAACATCAAAACTTC
25	Satt325	F	(ATT)19	GCGGGGTATTAAGGGAAAACAAAA	GCGTAAACGAACAATCACTTCATA
26	Sat_234	F	(AT)22	GCGATGCGTTTAATAAGTTTTGAAAAATGCC	GCGGAAACCATCCTTATATGTCAATTGCTCA
27	Satt610	G	(ATT)12	CCCTCCGCAAGCAATAATTAATCT	GCGGAATGCTTCCATTTTAT
28	Satt138	G	(ATT)47	GACATTTTTCCACGGATATTGAAT	AACGGGCGATTTATGGCTAT
29	Satt191	G	(ATT)18	CGCGATCATGTCTCTG	GGGAGTTGGTGTTTTCTTGTG
30	Satt314	H	(ATT)13	GCGGAGATTGGAACCTACTCATTC	GCGGGGACCAAAAATTCAAAA
31	Sat_180	H	(AT)23	GATCTAGGGCAAACAAGGTT	CTCGCTCTTCGCAACATA
32	Satt419	I	(ATT)22	GCGCCATGACGCCAGCTTGCTTTTCC	GCGCCATTATATACAAATTACGATCAT
33	Sct_189	I	(CT)17	CTTTTCCTGGCAATGAT	AAAATCGCAAAACCTTAGT
34	Sat_361	J	(AT)19	GCGTTAGATTTCCTTAGAATACATTGCTTCC	GCGTTGACACTCATGATGTTATCTTACACC
35	Sat_144	J	(AT)24(GA)11	GCGCCCTCTTCATTTCCCTTGGTT	GCGCCCAATATCTTTGGGATAAAT
36	Satt326	K	(ATT)18	AGATTCTCCTTTGCTTCTTAGT	GTTAGTTCACCTTCCAGTATTTGA
37	Satt588	K	(ATT)18(AT)10(CT)14	GCTGCATATCCACTCTCATTGACT	GAGCCAAAACCAAAGTGAAGAAC
38	Sat_340	L	(AT)31	GCGATGGAGCCGTGCTTATTCAC	GCGGCTTTCCCCTTTTAACTCTGAG
39	Satt373	L	(ATT)21	TCCGCGAGATAAATTCGTAAAAT	GGCCAGATACCCAAGTTGTACTTGT
40	Satt150	M	(ATT)20	AAGCTTGAGGTTATTCGAAAATGAC	TGCCATCAGGTTGTGTAAGTGT
41	Satt308	M	(ATT)21	GCGTTAAGGTTGGCAGGGTGGAAGTG	GCGCAGCTTTATACAAAAATCAACAA
42	Satt152	N	(ATT)14	GCGCTATTCCTATCACAACACA	TAGGGTTGTCACTGTTTTGTTCTTA
43	Sat_275	N	(AT)24	GCGCGCTGGCAATTATTCAAAACTTAACGAT	GCGAAGGCTACGGTGAATAGAAAGGAC
44	Sat_306	N	(AT)22(GT)8	GCGGTTCTGGGTTCTCTGTTTTCTTA	GCGTGATCCCACTATTTTAATTATACATTTT
45	Satt492	O	(ATT)15	GTATCGTTCGCGTCTTGAGTC	GCAGCGGTGTAGTTCGTTCTTTCT
46	Sat_282	O	(AT)21	GCGTCCCGATGATTCTTGGATCTA	GCGCGATTCTTGCCACTGTATTT