Received: 23 Dec 2021 | Revised: 15 Jan 2022 | Accepted: 28 Jan 2022

Electronic Journal of Plant Breeding

Research Article

Characterization of soybean genotypes based on morphological and molecular markers

Niharika Saharia* and Ramendra Nath Sarma

Department of Plant Breeding and Genetics, Assam Agricultural University, Jorhat-785013
***E-Mail:** niharikasaharia10@gmail.com

Abstract
The present investigation was carried out with 38 soybean genotypes during *Kharif*, 2020. The variability analysis was done using 10 quantitative and 11 qualitative characters. Twenty four SSR markers were used to study genetic relationships among the genotypes based on Jaccard's coefficient of similarity. The analysis of variance revealed that there were significant variations for yield and yield attributing characters. Plant height, the branches per plant, the pods per plant and seed yield showed the highest coefficients of variability along with high heritability with a genetic advance as percentage of mean. Correlation and path coefficient analyses identified plant height, the branches per plant and the pods per plant traits identified as important traits for yield improvement in soybean. Euclidean distance based on qualitative characters indicated the genotypes DLSb-2 and JS 22-14 are most divergent, while quantitative characters based analysis revealed that the genotypes NRC 128 and TS20-5 are potential for use in breeding. Out of 24 soybean specific SSR primers, 19 were found to be polymorphic. The number of SSR allele per locus ranged from one to three with an average of 1.4 alleles per locus. DNA marker analysis revealed a range of diversity in the experimental materials with few potential markers for diversity analysis due to their high PIC values.

Keywords: Euclidean distance, Genetic variability, Path coefficient analysis, Soybean, SSR marker

INTRODUCTION

Soybean [*Glycine max* (L.) Merrill] belonging to family 'Leguminosae' and subfamily 'Papilionoide' is an important grain legume crop. It is a multipurpose crop which is also referred to as "Golden jewel" or "King of Beans" or "Miracle crop of 21st century". Although it is a leguminous crop, it has gain popularity as an oilseed. In India, soybean was reported to be grown for ages under different vernacular names like Bhat, Bhatwan, Garikuley, Khajuwa or Kalitur in Garhwal hills of Uttarakhand, Kumaon hill and some places of Central India (Watt, 1890). It is a rich source of dietary protein that is consumed as edamame, soy milk, soya chunks, tofu, tempeh etc. (Rizzo and Baroni, 2018). The grain contains 36-56% high-quality protein and approximately 17-20% oil (Grieshop and Fahey, 2001; Van *et al.*, 2009). The improvement of soybean has been carried out through both conventional breeding methods and by utilizing biotechnological tools. Like other crops, loss of genetic diversity among the primitive soybean cultivars occurred during domestication and due to targeted breeding which is evident from the pedigree and diversity analyses conducted in the major soybean growing countries like the USA (Delannay *et al.*, 1983), Brazil (Wysmierski and Vello, 2013), India (Bhardwaj *et al.*, 2002), Japan (Zhou *et al.*, 2000). Thus, there is a need to enhance the genetic potential of soybean for sustained and long term improvements both in yield and quality parameters.

The information of variation present in a population is a pre-requisite for their efficient utilization in a breeding programme (Govindaraj *et al.*, 2015). Traditionally, the diversity is assessed through common morphological

traits (Cox *et al.*, 1985; Ariyo, 2004), but such assessments have many limitations. DNA based marker like, simple sequence repeats (SSR) is becoming increasingly popular in cultivar identification, and diversity analysis with great precision (Mondini *et al.*, 2009).

Since, adequate knowledge on genetic diversity will provide an opportunity for plant breeders to develop improved cultivars with desirable characteristics, thus the evaluation of genetic variability; heritability and correlation are of utmost importance. Therefore, the present investigation was carried out to study pattern of genetic variability and diversity using morphological characters and SSR markers in a set of 38 diverse soybean genotypes.

MATERIALS AND METHODS

The present investigation was carried out with 38 soybean genotypes which were received from All India Coordinated Research Project (AICRP) on Soybean, Jorhat centre during *Kharif*, 2020. These genotypes were selected based on a preliminary observation (data not shown) which reflected high variation in some morphological traits. In the field experiment, the data on eleven qualitative traits and ten quantitative traits were recorded altogether. The quantitative traits measured were plant height (cm), days to 50% flowering and maturity, branched per plant (number), pods per plant (number), seeds per pod (number), seed yield per plant (g), 100 seed weight (g) and protein and oil content (%). The flowering traits, maturity and qualitative traits were recorded on a plot basis. The other traits were recorded on plant basis, while protein and oil content was measured using of NI Grain Analyzer (NIR spectrometer model 6500, FOSS). The qualitative traits measured were plant growth type, shape of lateral leaflets, intensity of green colour, hilum colour, seed coat colour, pod colour after maturity, flower colour, pubescent colour, pubescent type, seed coat lustre and pubescent density.

Nature of variability on qualitative characters was analyzed based on frequency distribution. Quantitative traits were analyzed by Analysis of variance (Fisher's Method). The phenotypic, genotypic and environmental variances, coefficients of variation, heritability in broad sense and genetic advance as percentage of mean, phenotypic and genotypic correlation coefficients along with path coefficient analysis were calculated according to Singh and Choudhury (1985) using Variability package Ver. 0.1 in R platform (Anon, 2020). Euclidean distance for both qualitative and quantitative character was measured and relationship among genotypes were represented as dendrogram using UPGMA based clustering method in NTSYS-PC version 2.2 (Rohlf, 2000)

For molecular analysis, 24 SSR primers were used to asses fingerprint pattern for estimation of genetic relatedness. DNA was extracted using the CTAB protocol (Doyle and Doyle, 1990). Total 24 SSR markers were selected from initial screening with 10 genotypes. PCR reaction was carried out in a total volume of 10 µl containing 1.5µl genomic DNA (10 ng/µl), 1 µl of 10X buffer, 0.8 µl of 25 mM MgCl2, 0.2 µl of 10 mM dNTPs, 0.5µl of each primer (10 nmol), 0.2 µl of Taq DNA polymerase with volume make-up with distilled water. PCR reactions were carried out in a thermocycler (PCR Gene AMP® 2400, Applied Biosystems, USA) using the following cycling parameters: with initial denaturation 94 °C for 5 min., followed by 36 cycles of 94 °C for 30 s, 45–50 °C for 45 s , 72 °C for1min. and a final extension at 72 °C for 10 min. The amplified products were separated on 3 % agarose gels and detected by ethidium bromide staining and the gel pictures were recorded using gel documentation System (VILBER, France). Allele sizes were estimated in comparison with 100 bp DNA ladder. Amplified fragments were scored as binary data, *i.e.* presence as 1 and absence as 0.The polymorphic information content (PIC) values for each primer were calculated (Anderson *et al.*, 1993).Genetic relationships among individuals were quantified by the Jaccard's coefficient similarity. Graphical representation of similarity among the genotypes was depicted in dendrogram drawn using UPGMA method in NTSYS-PC version 2.2 (Rohlf, 2000).

RESULTS AND DISCUSSION

The importance of qualitative traits to characterize genetic divergence of soybean genotypes was highlighted by Chavez *et al.* (2017) and Dhaka *et al.* (2020). Three qualitative characters under study *viz.* seed coat colour, flower colour and pubescent colour showed only two classes (bi-morphic) while the rest were polymorphic. Among all the 11 qualitative characters studied, high morphological variability, in terms of more classes, was observed for hilum colour, pubescent density and pubescent colour with four classes each, while plant growth type, the shape of the leaflet, intensity of green colour, pod colour after maturity and the seed coat lustre with three classes each. The frequency distribution of different forms of qualitative traits are shown in **Fig.1 (a-k)**. This result suggested presence of high variation among the genotypes indicating their importance in varietal discrimination and diversity analysis (Gawande *et al.*, 2001).

The ANOVA revealed a significant variation among all the ten quantitative characters indicating the presence of variability which might provide scope for their exploitation in breeding programme (Karnwal and Singh, 2009; Baraskar *et al.,* 2014) (**Table 1**). High GCV and PCV were recorded for plant height, the branches per plant and the pods per plant and seed yield per plant (Baraskar *et al.,* 2014; Chandrawat *et al.,* 2017) (**Table 2**). It is expected that the characters exhibiting high GCV would exhibit a response to selection in a positive direction. High heritability along with high genetic advance (as % of mean) was recorded for the plant height, the branches

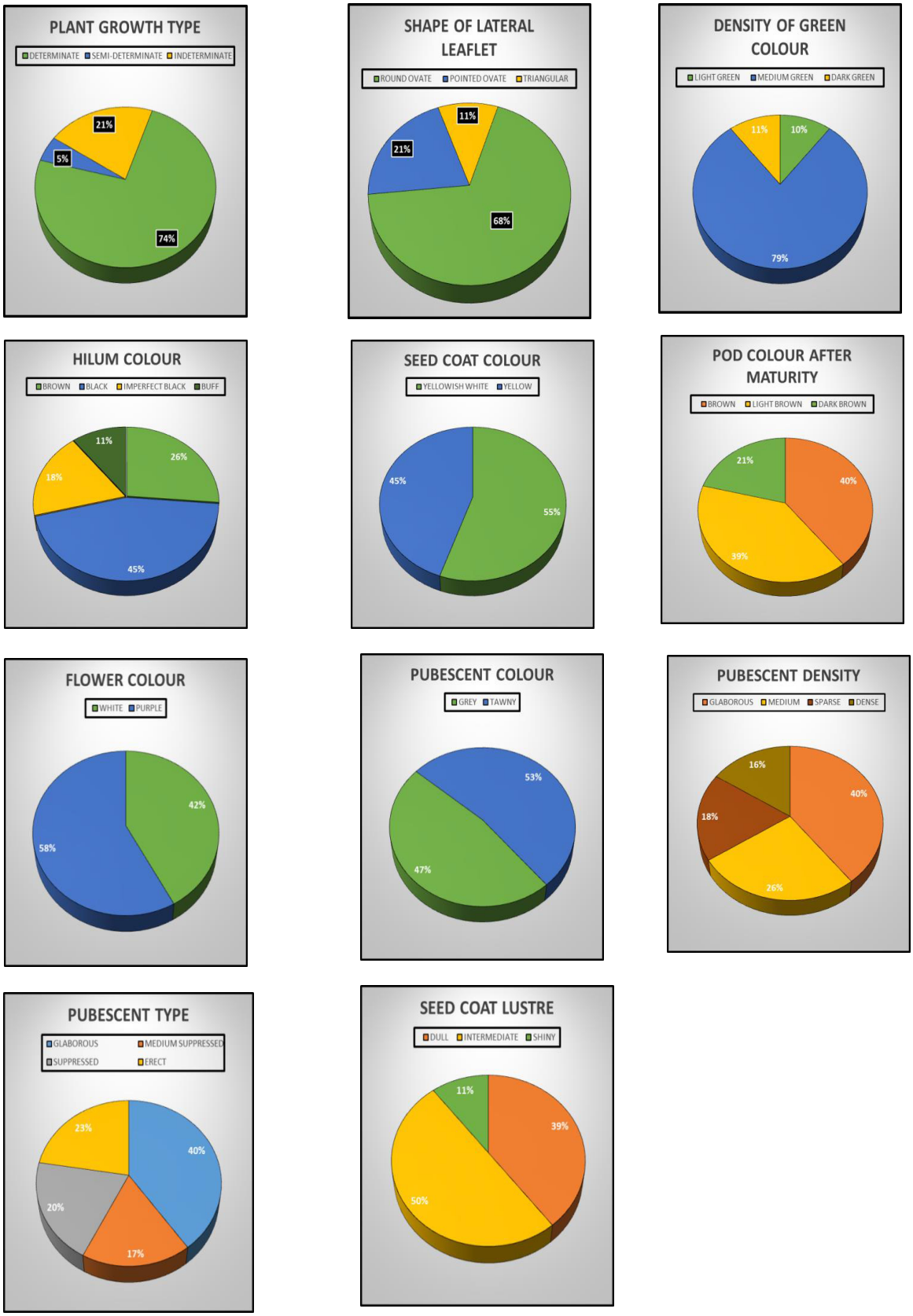

Fig. 1. Distribution of different qualitative traits in 38 soybean genotypes

Table 1. ANOVA for 10 quantitative characters in 38 soybean genotypes

S. No.	Characters	Source of variation		
		Replication	Treatment	Error
	Degree of freedom (df)	(2)	(37)	(74)
1.	Plant height	118.080	477.752**	17.489
2.	Days to 50% flowering	2.904	18.120**	0.687
3.	Days to maturity	1.482	254.174**	2.059
4.	Number of branches/plant	13.167	6.318**	2.077
5.	Number of pods/plant	27.798	377.979**	114.672
6.	Number of seeds/pod	0.061	0.455**	0.197
7.	Seed yield/plant	8.980	62.016**	8.109
8.	Seed weight	2.167	3.323**	0.644
9.	Protein content	0.076	14.106**	0.010
10.	Oil content	0.116	5.571**	0.017

**1% significant level

Table 2. Genetic parameters of variations for different characters in soybean

	PH	DFL	DM	NBP	NPP	NSP	SW	PRO	OIL	SYP
GCV %	21.31	5.892	9.094	20.512	20.374	10.564	7.381	5.554	5.997	22.356
PCV %	22.492	6.231	9.205	22.803	23.348	19.125	9.682	5.560	6.025	26.932
H^2(BroadSense) %	0.898	0.894	0.976	0.899	0.872	0.305	0.581	0.998	0.991	0.689
GAM@ 5%	41.592	11.478	18.508	27.791	20.853	12.021	11.591	10.429	10.297	38.231

PH- plant height, DFL: days to 50% flowering, DM= Days to maturity, NBP= number of branches per plant, NPP=number of pods per plant, NSP= number of seeds per pod, SYP= seed yield per plant, SW= seed weight, PRO= protein content, OIL=oil content

per plant, the pods per plant and the seed yield per plant indicating the preponderance of additive gene action and phenotypic selection for the traits could be useful (Singh *et al.*, 2000; Jain and Ramgiry, 2000). Low genetic advance coupled with high heritability was recorded for protein and oil content indicating that these traits might be governed by genes with non-additive gene action. Thus, it has been suggested that improvement of these traits either through heterosis breeding or population improvement (Ramteke *et al.,* 2010). Combining all variability parameters, the plant height, the branches per plant and the pods per plant could be useful in selection programme among the genotypes.

Mean performance of the 38 genotypes revealed the potentiality of the test genotypes for different yield attributing traits (**Table 3**). NRC-128 had high *per se* performance for seed yield, pods per plant and oil content. JS 22-11 had the highest seed weight and is the tallest among all the genotypes. Though ASb-9 flowered the earliest, the genotype DS 3105 matured early with the highest seeds per pod. The genotype with the highest protein content was NRC-109. Thus, it can be concluded that genotype NRC-128, JS 22-11, ASb-9, DS 3105 and NRC-109 were the best genotypes based on *per se* performances for yield components, while none of the genotype showed desirable *per se* performance for the maximum characters.

It was observed that the genotypic correlation coefficients were higher than their corresponding phenotypic values for all the traits studied indicating a greater contribution of genotypic factors and therefore genotypic correlation coefficients were taken in to account for all analyses **(Table 4).** The seed yield per plant showed a positive correlation with the plant height, the branches per plant and the pods per plant. The results were in accordance with the finding of Hatam (2001). The positive correlation of seed yield per plant with pods per plant was reported by Machikowa and Laosuwan,(2011); with branches per plant and pods per plant by Iqbal *et al.*(2003) and Malik *et al.*(2006); with plant height and the pods per plant by Baig *et al.* (2017). This indicated that seed yield per plant could be increased by selecting for tall plants with concomitant increase in branches per plant and pods per plant (Balla and Ibrahim, 2017).

The path coefficient analysis using genotypic correlation coefficients, with seed yield per plant as dependant variable, showed that highest positive direct effect by

Table 3. Mean Performance of 38 soybean genotypes based on quantitative characters

Genotypes	PH	DFL	DM	NBP	NPP	NSP	SW	PRO	OIL	SYP
DSb-38	51.533	40.333	95.000	8.000	59.333	3.000	12.333	39.733	22.800	22.833
DS 3105	54.167	40.667	76.000	9.000	57.333	2.000	12.767	38.667	22.467	23.300
CAUM 52	62.400	41.667	96.000	10.000	73.667	2.333	12.000	40.067	23.800	23.167
JS 22-11	85.233	41.333	105.000	9.000	52.667	2.333	14.233	40.433	21.700	20.440
DLSb-2	53.867	43.667	93.667	11.667	62.667	2.667	11.333	37.433	23.167	23.000
RVSM 2012-11	64.167	42.000	93.667	6.667	64.667	2.667	14.133	41.033	21.133	21.733
RSC 11-39	83.233	40.667	103.000	8.333	58.000	2.667	11.800	39.700	21.400	20.933
AS-15	73.233	43.000	106.000	9.667	88.000	3.000	13.867	40.533	21.100	24.853
PS 1664	46.567	40.333	102.667	9.667	59.000	2.333	14.100	40.267	21.667	15.433
HIMSO 1691	48.467	36.667	100.333	6.667	82.000	3.000	12.767	35.833	24.667	22.517
JS 22-14	53.633	40.667	97.000	5.667	49.000	3.000	13.900	35.567	23.467	20.733
DS 3144	53.867	42.000	112.000	7.333	53.667	2.667	14.067	41.067	23.300	22.167
DLSb-1	61.567	43.333	113.667	8.333	69.000	2.667	13.333	37.700	22.500	22.533
NRC 128	70.600	40.667	102.667	8.333	94.333	2.667	13.067	33.133	25.567	26.167
VLS 101	50.100	42.000	88.000	6.000	74.333	2.333	12.700	39.133	23.933	21.933
RSC 11-35	70.800	43.333	112.000	8.000	53.333	3.000	13.000	39.200	22.667	22.633
PS 1661	63.600	37.333	102.000	7.333	71.000	2.667	12.100	40.067	21.067	18.267
HIMSO 1692	53.400	38.333	97.667	7.667	61.667	2.333	11.533	42.267	22.800	21.967
JS 20-116	78.533	43.000	101.667	8.333	66.000	3.000	13.167	39.467	22.000	18.783
RVS 2012-10	49.100	42.333	110.667	8.333	55.333	3.333	11.667	41.233	20.967	20.100
PS 1670	66.367	37.667	104.667	10.000	61.333	3.000	13.267	40.067	21.833	19.583
NRC 109	74.967	38.333	101.333	9.667	59.333	2.333	10.933	42.833	20.133	21.700
MAUS 806	50.033	41.333	102.667	7.667	71.000	2.333	13.967	39.667	21.400	19.743
RVS 2011-10	56.233	37.667	115.000	7.667	58.667	3.333	13.800	40.167	21.267	16.100
MAUS 768	51.233	40.667	108.000	6.667	46.000	3.000	13.367	39.567	22.767	10.633
ASb 26	43.167	38.333	96.333	6.000	48.000	3.000	11.933	39.267	22.267	9.760
ASb 9	52.933	36.000	86.333	9.000	58.667	2.667	14.000	39.033	21.667	14.320
RKS 113	64.033	44.667	86.000	9.333	50.333	3.000	11.133	35.667	24.300	19.703
MACS 1701	49.133	37.667	111.000	9.000	62.333	3.667	11.233	37.233	23.733	19.533
KDS 1096	58.800	41.667	86.667	11.333	59.333	3.000	14.033	35.033	23.033	22.167
MACS 1691	54.300	42.000	88.000	9.000	53.333	3.333	13.833	37.300	24.267	11.433
KDS 114	54.133	43.000	112.333	9.333	71.333	3.000	12.567	35.267	24.967	22.333
BAUS 96-17	71.167	44.000	105.667	5.667	51.000	2.667	13.733	40.200	24.300	9.367
BAUS 31-17	76.100	38.333	107.000	8.333	60.000	2.000	11.267	41.567	20.433	16.533
TS 20-5	44.600	37.667	101.667	6.333	46.667	2.333	12.633	41.100	22.033	9.283
SL 1212	26.967	45.667	97.667	7.333	45.333	3.333	11.633	37.000	24.933	15.300
SL 1250	41.900	43.667	114.667	7.667	52.000	3.000	13.800	41.267	23.533	15.000
DS 1312	44.600	43.000	97.000	7.333	56.000	3.000	11.567	39.367	23.133	14.493
Mean	58.125	40.912	100.807	8.193	60.939	2.781	12.804	39.030	22.689	18.960
C.V.	7.194	2.026	1.423	12.589	12.573	10.943	6.265	0.252	0.579	10.018
C.D. 5%	6.804	1.352	2.333	2.634	17.422	0.727	1.312	0.168	0.214	4.633

PH- plant height (cm), DFL: days to 50% flowering, DM= Days to maturity, NBP= number of branches per plant, NPP=number of pods per plant, NSP= number of seeds per pod, SYP= seed yield per plant (g), SW= seed weight (g), PRO= protein content (%), OIL=oil content (%)

Table 4. Genotypic (above diagonal) and phenotypic (below diagonal) correlation coefficient

Traits	PH	DFL	DM	NBP	NPP	NSP	SW	PRO	OIL	SYP
PH	1.000	-0.012	0.133	0.347*	0.344*	-0.383*	0.079	0.136	-0.364*	0.454**
DFL	-0.008	1.000	0.017	0.093	-0.144	0.254	0.071	-0.197	0.390*	0.144
DM	0.123	0.010	1.000	-0.179	0.051	0.324*	0.107	0.258	-0.142	-0.057
NBP	0.175	0.092	-0.124	1.000	0.339*	-0.066	-0.148	-0.204	-0.135	0.553**
NPP	0.188	-0.061	0.042	0.115	1.000	-0.266	0.062	-0.314*	0.120	0.707**
NSP	-0.230	0.077	0.204	0.006	-0.059	1.000	0.024	-0.469**	0.415**	-0.244
SW	0.070	0.060	0.076	-0.097	0.048	-0.035	1.000	-0.017	-0.033	-0.069
PRO	0.129	-0.188	0.254	-0.130	-0.201	-0.272	-0.012	1.000	-0.701**	-0.221
OIL	-0.343*	0.365*	-0.141	-0.084	0.080	0.227	-0.021	-0.696**	1.000	0.043
SYP	0.316*	0.105	-0.030	0.277	0.557**	-0.068	-0.086	-0.267	0.049	1.000

** & * Significant at 1 % & 5 % level, respectively
PH- plant height, DFL: days to 50% flowering, DM= Days to maturity, NBP= number of branches per plant, NPP=number of pods per plant, NSP= number of seeds per pod, SYP= seed yield per plant, SW= seed weight, PRO= protein content, OIL=oil content

pods per plant followed by the branches per plant and days to flowering, respectively on seed yield (Machikowa and Laoswan, 2011) **(Table 5).** Also, Iqbal *et al.* (2003) reported a high direct effect for the branches per plant and pods per plant. These results suggested that a slight increase in one of the above traits may directly contribute to seed yield because of their position associations with seed yield per plant. The highest indirect effect on seed yield per plant was observed for plant height *via* the pods per plant which was in accordance with Iqbal *et al.*(2003); Balla and Ibrahim (2017). Also, high indirect effect was shown by the branches per plant via the pods per plant which was in agreement with Arshad *et al.* (2014). Therefore, the pods per plants could be used as a direct or indirect selection criterion in identification for higher-yielding genotypes, while at the same time given more branches per plants to the tall plant. This suggested that selection on the basis of the pods per plants would be efficient (Balla and Ibrahim, 2017).

Based on qualitative characters, the maximum Euclidean distance was observed between DLSb-2 and JS 22-14 (8.07) indicating the most diverse genotypes as they had dissimilarity for almost all the qualitative characters except for medium green colour leaves, purple flower and brown pods after maturity **(Fig.2).** The dendrogram grouped the genotypes into two major clusters; cluster A (31 genotypes) and cluster B (5 genotypes). Two solitary genotypes *viz.,* JS 22-11 and RVS 2011-10 were also observed. Cluster A was further subdivided into two sub-clusters, the sub-cluster I with 18 genotypes having determinate growth habit with round ovate leaves and sub-cluster II with 13 genotypes that had tawny pubescent and light brown coloured pods after maturity. The genotypes belonging to Cluster B showed indeterminate growth types. All of them had medium green coloured leaves and grey pubescent. The solitary genotype JS 22-11 was characterized by indeterminate growth type with pointed ovate leaves. Other solitary genotype RVS 2011-10 could be distinguished from others by its triangular light green colour leaves and white flower. This result indicated a great diversity among the soybean genotypes to be useful for germplasm curator in their maintenance and in plant variety protection. Importantly, some of the genotypes (JS 22-14, DLSb 2, NRC-128, NRC-109) can be used in breeding for development of high yielding variety.

Considering quantitative data, the maximum distance is found between the genotypes NRC 128 and TS20-5 (7.78) as their mean value for most of the characters were different except for days to maturity and pods per plant **(Fig.3).** Dendrogram constructed by UPGMA cluster analysis grouped the genotypes into two board clusters; cluster I (34 genotypes) and Cluster II (4 genotypes). The sub-cluster I of the main cluster I was characterized by the highest mean for plant height and protein content. Sub-cluster II of Cluster I had the highest mean for days to flowering. Sub-cluster III was characterized by the highest branches per plant and earliest in maturing. Main cluster II comprised of genotypes characterized by highest pods per plant, seeds per pod, seed yield, and oil content. These results indicated that a careful selection of parents from different cluster might contribute in the evolution of desirable segregants.

The molecular diversity assessed by twenty-four SSR markers revealed that nineteen of them were polymorphic and hence used for further analysis **(Table 6).** The number of amplified products ranged from one to three, which were considered as alleles. Maximum numbers of alleles were amplified by the primer Satt292, Sat_172 and Sat_137. Seven primers amplified two alleles and eleven primers amplified only one allele with an average of 1.4 alleles per locus. High average alleles per locus have been reported by several workers. The study by Diwan and Cregan (1997) had reported as high as 10.1

Table 5. Direct and Indirect effects of various characters on seed yield at genotypic level

Traits	PH	DFL	DM	NBP	NPP	NSP	SW	PRO	OIL
PH	-0.047	0.001	-0.006	-0.016	-0.016	0.018	-0.004	-0.006	0.017
DFL	-0.003	0.283	0.005	0.026	-0.041	0.072	0.020	-0.056	0.111
DM	0.010	0.001	0.073	-0.013	0.004	0.024	0.008	0.019	-0.010
NBP	0.094	0.025	-0.048	0.270	0.091	-0.018	-0.040	-0.055	-0.037
NPP	0.198	-0.083	0.029	0.195	0.576	-0.153	0.036	-0.181	0.069
NSP	0.081	-0.054	-0.069	0.014	0.056	-0.212	-0.005	0.099	-0.088
SW	-0.009	-0.008	-0.012	0.016	-0.007	-0.003	-0.109	0.002	0.004
PRO	-0.027	0.039	-0.051	0.040	0.062	0.093	0.003	-0.198	0.139
OIL	0.057	-0.061	0.022	0.021	-0.019	-0.064	0.005	0.109	-0.155
SYP(r)	0.454**	0.144	-0.057	0.553**	0.707**	-0.244	-0.086	-0.267	0.049

RESIDUAL EFFECT = 0.56

PH- plant height, DFL: days to 50% flowering, DM= Days to maturity, NBP= number of branches per plant, NPP=number of pods per plant, NSP= number of seeds per pod, SYP= seed yield per plant, SW= seed weight, PRO= protein content, OIL=oil content

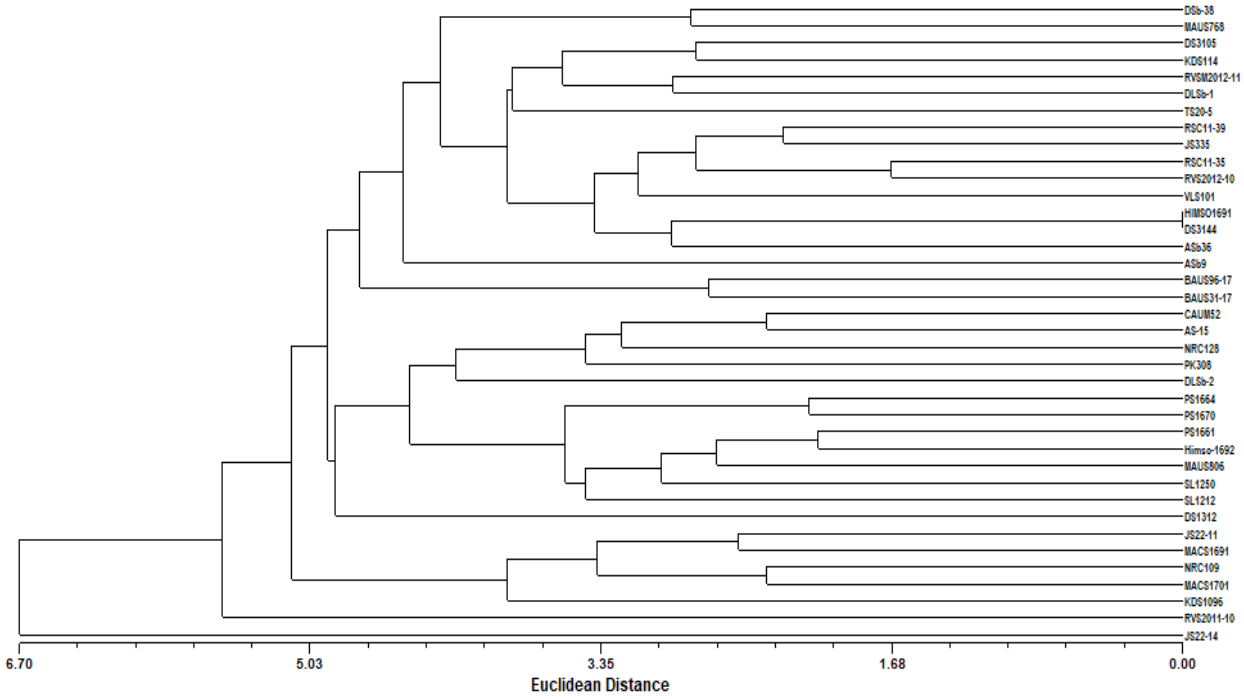

Fig. 2. Euclidean distance based dendrogram for qualitative characters

alleles per locus in ancestral soybean genotypes. Other similar studies was reported by Wang *et al*. (2008) and Yoon *et al*. (2009). The higher average alleles per locus observed in those studies might be due to the reason that a higher number of accessions from diverse geographical locations were used as compared to the present study. So, no parallelism of the current investigation can be drawn with those studies in terms of level of allelic diversity. No unique allele distinguishing a particular genotype was found.

The PIC value is the measure of allelic diversity of SSR ranged from 0.10 to 0.79 which was in close agreement with Kumawat *et al.* (2015). The maximum PIC was revealed by primer Sat_172, while minimum PIC among polymorphic markers was shown by primer Satt416. The average value of PIC was found to be 0.37 which is in conformity with Hipparagi *et al.* (2017). The PIC values revealed that the primers with higher PIC values are informative and useful for further genetic diversity studies. Markers with PIC values greater than 0.50 and high

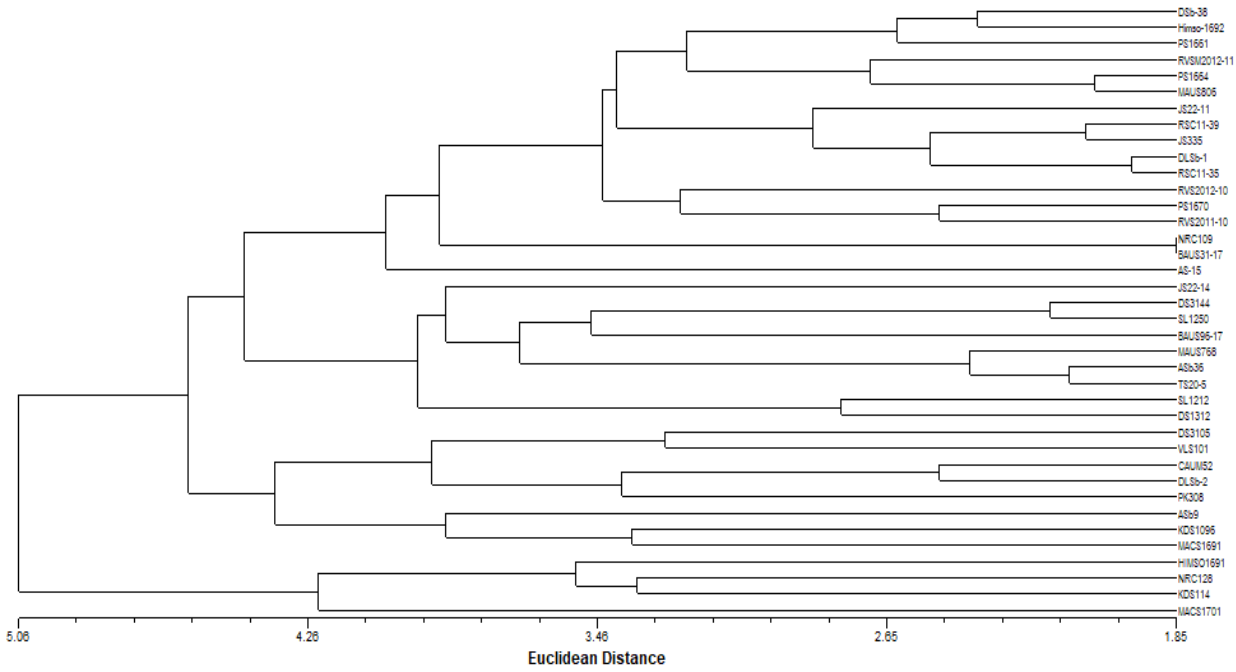

Fig. 3. Euclidean distance based dendrogram for quantative characters

Table 6. DNA amplification profile and PIC values of 24 SSR markers

S.No.	Markers	Amplicon Size (bp)	Number of alleles amplified	Polymorphic allele	Polymorphic frequency (%)	PIC
1	Satt701	250	1	1	100	0.57
2	Satt292	400,280,200	3	3	100	0.55
3	Sat_137	350,290,100	3	2	66	0.12
4	Satt277	300,200	2	2	100	0.53
5	Satt306	250	1	1	100	0.25
6	Sat_316	250,110	2	2	100	0.35
7	Satt243	200	1	1	100	0.10
8	Satt665	290	1	1	100	0.25
9	Sat_172	230,200,100	3	3	100	0.79
10	Satt152	300,250	2	2	100	0.57
11	Sat_225	220,100	2	2	100	0.45
12	Sat_140	200,110	2	1	50	0.17
13	Sat_147	300	1	0	0	0.00
14	Sat_330	290,110	2	1	50	0.47
15	Sat_217	280,150	2	1	50	0.17
16	Satt424	250,120	2	2	100	0.59
17	Satt260	300	1	1	100	0.72
18	Sat_309	300	1	0	0	0.00
19	Satt416	350	1	1	100	0.25
20	Sat388	300	1	0	0	0.00
21	Satt220	280,200	2	1	50	0.28
22	Sat_305	340	1	0	0	0.00
23	Sat_216	310,280	2	1	50	0.27
24	Satt291	220	1	0	0	0.00

allele numbers were most informative for genetic studies (Moniruzzaman *et al.*, 2019). Accordingly, the primers Satt701, Satt292, Satt27, Sat_172, Satt152, Satt424 and Satt260 were useful in discrimination of soybean genotypes.

Jaccard's coefficient of similarity ranged from 0.52 (between BAUS 31-17 and DLSb-1) to 0.97 (between CAUM 52 and RK-113) with an average value of 0.72. It indicated the maximum genetic diversity between BAUS 31-17 and DLSb-1 and the highest genetic similarity between CAUM 52 and RK-113. A similar coefficient of similarity range was reported by Sneller *et al.* (1997) and Bisen *et al.* (2015). However, the result obtained is in conflict with the observation of Priolli *et al.* (2002) in which the genetic dissimilarity coefficients were relatively high with a mean of 0.46.

Based on the dendrogram generated through the UPGMA method, two major clusters (A and B) were observed and cluster B was the largest cluster with two sub-clusters and a total of 21 genotypes **(Fig.4).** A low level of genetic diversity might be attributed to narrow range of genetic variability among the materials with their limited number under study. Due to lack of pedigree information of the genotypes under study no further analysis of close relatedness of the genotypes could be discussed.

The present investigation revealed a high range of variation among the genotypes. Variability, correlation and path coefficient analysis identified that the traits plant height, branches per plant and pods per plant are the most important characters warranting due importance in soybean improvement. The *per se* performance and diversity analyses identified NRC-128, JS 22-11, DLSb-2, NRC 109, JS 22-14, TS 20-5, DLSb-1 and BAUS31-17 as potential parents to use in hybridization for obtaining desirable segregants. The lack of correspondence between the clustering patterns based on qualitative and quantitative traits using Euclidean distance was observed and this might be due to different gene system for two types of traits which could not be confirmed in this study. Moreover, the lower agreement between SSR and morphological distances may be due to the fact that the variation observed at SSR level might have not been expressed at phenotypic level. Limited numbers of markers were analyzed in this study and the distribution of the markers studied was not enough to cover the whole genome of soybean. Also, there was a considerable effect of environment on morphological traits, thus there is less agreement between the diversity pattern of morphological traits and molecular markers. The primers Satt701, Satt292, Satt27, Sat_172, Satt152, Satt424 and Satt260 were useful to study diversity and mapping.

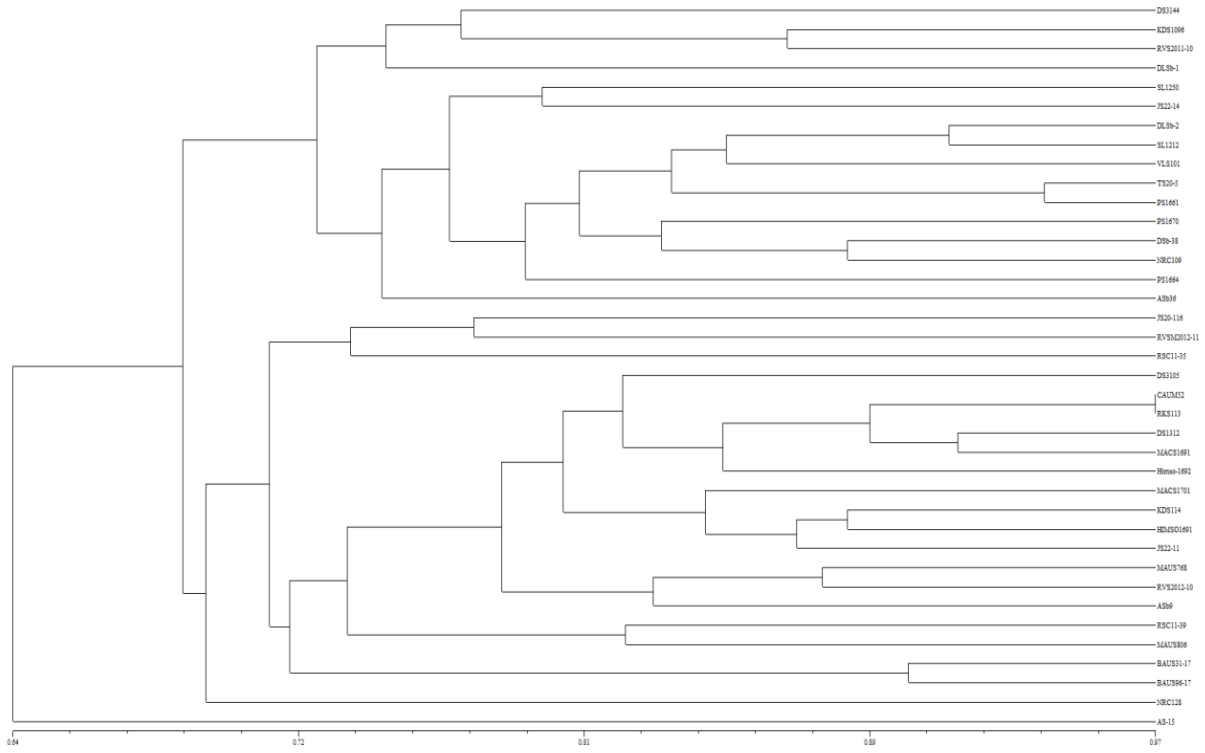

Fig. 4. Jaccard's coefficient based dendrogram for SSR markers

REFERENCES

Anderson, A., Churchill, G. A., Autrique, J. E., Tanksley, S. D. and Sorrells, M. E. 1993. Optimizing parental selection for genetic linkage maps. *Genome*, **36**(1): 3-7. [Cross Ref]

Anonymous 2020. .https://cran.r-project.org/web/packages/variability/index.html as visited on 1st November 2021.

Ariyo, O. J. 2004. Genetic diversity in West African okra (*Abelmoschuscaillei*) (*A. Chev*.) Stevels-Multivariate analysis of morphological and agronomic characteristics. *Genetic Resources and Crop Evolution*, **40**: 25-32. [Cross Ref]

Arshad, W., Zeeshan, M., Khan, M.I., Ali, S., Hussain ,M. and Rahman, S.U. 2014. Character association and causal effect analysis for yield and yield components among early maturing genotypes of soybean (*Glycine max* (L.) Merrill). *Journal of Renewable Agriculture*, **2**: 1-4. [Cross Ref]

Baig, K.S., Jadhav, P.P., Sarang, D.H. and Chandrawat, K.S. 2017. Correlation and path analysis studies in soybean (*Glycine max* (L.) Merrill.), *International Journal of Pure and Applied Bioscience*, **5**(4): 489-492. [Cross Ref]

Balla, M. Y. and Ibrahim, S. E. 2017. Genotypic correlation and path coefficient analysis of soybean (*Glycine max* (L.) Merr.) for yield and its components. *Agric Res Tech*, **7**(3): 1-5.

Baraskar, V. V., Kachhadia, V. H., VachhanI, J. H., Barad, H. R., Patel, M. B. and Darwankar, M. S. 2014. Genetic variability, heritability and genetic advance in soybean (*Glycine max* (L.) Merrill). *Electronic Journal of Plant Breeding*, **5**(4): 802-806.

Bhardwaj C. H., Satyavathi C. T., Tiwari S. P. and Karmakar P. O. 2002. Genetic base of soybean varieties released in India as revealed by coefficient of parentage. *Indian Journal of Agricultural Science*, **72**: 567-469.

Bisen, A., Khare, D., Nair, P. and Tripathi, N. 2015. SSR analysis of 38 genotypes of soybean (*Glycine Max* (L.)Merr.) genetic diversity in India. *Physiology and molecular biology of plants*, **21**(1): 109–115. [Cross Ref]

Chandrawat, K. S., Baig, K. S., Hashmi, S., Sarang, D. H., Kumar, A. and Dumai, P. K. 2017. Study on genetic variability, heritability and genetic advance in soybean. *Int. J. Pure App. Biosci*, **5**(1): 57-63. [Cross Ref]

Chavez, H. D., Borromeo, T., Borines, N. O. and Gentallan Jr, R. 2017. Phenotypic diversity of soybean (*Glycine max* (L.) Merr.) accessions in the Philippines for utilization. *Legume Research*, **40**(1): 9-15. [Cross Ref]

Cox, T.S., Kiang, Y., Gorman, M. and Rodgers, D.M. 1985. Relationship between coefficient of parentage and genetic similarity indices in the soybean. *Crop Science*, **25**: 529-532. [Cross Ref]

Delannay, X., Rodgers, D. M. and Palmer, R. G. 1983. Relative genetic contributions among ancestral lines to north American soybean cultivars1.*Crop Science*, **23**(5): 944-949. [Cross Ref]

Dhaka, N.S., Singh, K., Shukla, P. S, Jeena, A. S. and Kumar, S. 2020. Qualitative characterization of diverse germplasm of soybean (*Glycine max* (L.) Merrill. *International Journal of Chemical Studies*, **8**(6):866-872. [Cross Ref]

Diwan, N. and Cregan, P.B. 1997. Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. *TAG. Theoretical and applied genetic,***95**:723-733. [Cross Ref]

Doyle, J.J. and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. *Focus*, **12**: 13-15. [Cross Ref]

Gawande, P., Patil, S. and Joshi, S. 2001. Selection of superior germplasm lines for soybean based on morphological characters for hybridization. *Journal of Physiological Research*, **14**(2):125-130.

Govindaraj, M., Vetriventhan, M. and Srinivasan, M. 2015. Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. *Genetics research international*, Article ID 431487. [Cross Ref]

Grieshop, C. M. and Fahey, G. C, Jr. 2001. Comparison of quality characteristics of soybeans from Brazil, China, and the United States. *Journal of agricultural and food chemistry*, **49**(5): 2669–2673. [Cross Ref]

Hatam, M. 2001. Performance of AVRDC vegetable soybean germplasm under Peshawar valley conditions [Pakistan]. *Sarhad Journal of Agriculture*, **17**(1): 27-30.

Hipparagi, Y., Singh, R., Choudhury, D. R. and Gupta, V. 2017. Genetic diversity and population structure analysis of Kala bhat (*Glycine max* (L.) Merrill) genotypes using SSR markers. *Hereditas*, **154**(1): 1-11. [Cross Ref]

Iqbal, S., Mahmood, T., Ali, M., Anwar, M. and Sarwar, M. 2003. Path coefficient analysis in different genotypes of soybean (*Glycine max* (L) Merril). *Pakistan Journal of Biological Sciences* (Pakistan), **6**: 1085-1087. [Cross Ref]

Jain, P. K. and Ramgiry, S. R. 2000. Genetic variability of metric traits in Indian germplasm of soybean (*Glycine max* (L.) Merrill). *Advances in Plant Sciences*, **13**(1): 127-131.

Karnwal, M. K. and Singh, K. 2009. Studies on genetic variability, character association and path coefficient for seed yield and its contributing traits in soybean [*Glycine max* (L.) Merrill]. *Legume Research-An International Journal*, **32**(1): 70-73.

Kumawat, G., Singh, G., Gireesh, C., Shivakumar, M., Arya, M., Agarwal, D. K. and Husain, S. M. 2015. Molecular characterization and genetic diversity analysis of soybean (*Glycine max* (L.) Merr.) germplasm accessions in India. *Physiology and molecular biology of plants*, **21**(1):101–107. [Cross Ref]

Machikowa, T. and Laosuwan, P. 2011. Path coefficient analysis for yield of early maturing soybean. *Songklanakarin Journal of Science & Technology*, **33**(4).

Malik, M. F. A., Ashraf, M., Qureshi, A. S. and Ghafoor, A. 2006. Utilization of diverse germplasm for soybean yield improvement. *Asian Journal of Plant Sciences*, **5**(4): 663-667. [Cross Ref]

Mondini, L., Noorani, A. and Pagnotta, M. 2009. Assessing plant genetic diversity by molecular tools. *Diversity*, **1**(1): 19-35. [Cross Ref]

Moniruzzaman, M., Saiem, R. M., Emon, R. M., Haque, M. S., Saha, N. R.,Malek, M. A. and Khatun, K. 2019. Genetic diversity analysis of soybean genotypes using SSR markers for salinity tolerance. *Progressive Agriculture*, **30**(1): 1-9. [Cross Ref]

Priolli, R. H. G., Mendes-Junior, C. T., Arantes, N. E. and Contel, E. P. B. 2002. Characterization of Brazilian soybean cultivars using microsatellite markers. *Genetics and molecular biology*, **25**(2): 185-193. [Cross Ref]

Ramteke, R., Kumar, V., Murlidharan, P. and Agarwal, D. K. 2010. Study on genetic variability and traits interrelationship among released soybean varieties of India (*Glycine max* (L.) Merrill). *Electronic Journal of Plant Breeding*, **1**(6): 1483-1487.

Rizzo, G. and Baroni, L. 2018. Soy, Soy Foods and Their Role in Vegetarian Diets. *Nutrients*, **10**(1): 43. [Cross Ref]

Rohlf F. J. 2000. NTSYS-pc, Numerical Taxonomy and Multivariate Analysis System. Version 2.0, Exeter software, New York

Singh J., Parmar, R. P. and Yadav, H. S. 2000. Assessment of genetic variability and selection parameters in early generation of soybean. *Advances in Plant Sciences*, **13**(1): 227-232.

Singh, R.K. and Chaudhary, B.D. 1985. Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi, Ludhiana, India. pp. 39-78.

Sneller, C. H., Miles, J. W. and Hoyt, J. M. 1997. Agronomic performance of soybean plant introductions and their genetic similarity to elite lines. *Crop Science*, **37**(5): 1595-1600. [Cross Ref]

Van Ee J. H. 2009. Soy constituents: modes of action in low-density lipoprotein management. *Nutrition reviews*, **67**(4): 222–234. [Cross Ref]

Wang, L. X., Guan, R. X., Li, Y. H., Lin, F. Y., Luan, W. J., Li, W. and Qiu, L. J. 2008. Genetic diversity of Chinese spring soybean germplasm revealed by SSR markers. *Plant Breeding*, **127**(1), 56-61.

Watt, G. 1890. A Dictionary of the Economic Products of India. Volume III Cosmo Publication, Delhi, India, pp. 509-511. (Second reprint 1972). https://archive.org/details/in.ernet.dli.2015.109072/page/n155/mode/2up.

Wysmierski, P. T. and Vello, N. A. 2013.The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. *Genetics and molecular biology*, **36**(4): 547–555. [Cross Ref]

Yoon, M. S., Lee, J., Kim, C. Y., Kang, J. H., Cho, E. G. and Baek, H. J. 2009. DNA profiling and genetic diversity of Korean soybean (*Glycine max* (L.) Merrill) landraces by SSR markers. *Euphytica*, **165**(1): 69-77. [Cross Ref]

Zhou,X., Carter,T.E., Jr.,Cui, Z., Miyazaki, S. and Burton, J.W. 2000. Genetic base of Japanese soybean cultivars released during 1950 to 1988. *Crop Science*, **40(6)**: 1794–1802. [Cross Ref]