Canonical Discriminant Analysis for the Assessment of Genetic Variation in Soybean (*Glycine max* (L.) Merrill)

Bhawana Sharma*[2] and Brijvir Singh[1]

[1]*Department of Genetics and Plant Breeding, College of Agriculture, GB Pant University of Agriculture & Technology, Pantnagar-263145, Uttarakhand*

[2]*IRRI - India office, based at Indira Gandhi Krishi Viswavidyalaya, Raipur-492006, Chhattisgarh*

(Received: 29 January 2011; Revised: 06 January 2012; Accepted: 12 June 2012)

For improvement of soybean, it is necessary to gain insight into the magnitude of variability present in the species. The study was conducted to assess sources of genetic and phenotypic variability in 62 soybean varieties. Seventeen morphological, agronomical and quality characters were measured. The multivariate data set was analyzed by Canonical Discriminant Analysis (CDA) in combination with a clustering procedure. In this analysis, the first four canonical varieties were significant and, number of pods/plant was the most differentiating traits among the varieties. The canonical roots clustered the varieties into eighteen groups on the basis of the differentiating traits. CDA effectively analyzed the genetic variation and identified the traits that could better describe the variation among soybean genotypes. Cluster analysis was successful in differentiating the varieties into similar subgroups on the basis of the measured traits.

Key Words: Canonical discriminant analysis, Cluster, Soybean

Introduction

Soybean (*Glycine max* (L.) Merrill) is one of the most important crops in the world today by virtue of its richness in balanced protein and oil. However, overall improvement in yield, oil and protein content in soybean remains a perpetual task to be accomplished by the plant breeders. This can be achieved either by selecting superior genotypes or to be utilized as parents for the development of future cultivars through hybridization, from the variability existing in the available genepool. However, in Indian soybean varieties, a narrow genetic base has been observed (Chung and Singh, 2008; Pushpendra *et al.,* 2008), whereas, now-a-days breeding programmes use soybean improved cultivars in hybridization to create genetic variability. The use of related wild species is at minimum bringing about a very low impact on broadening of genetic base in soybean. Hence knowledge of genetic divergence in the available cultivars of soybean has an immense importance and in tune with immediate need in the selection of parents to be used in hybridization programme for obtaining desirable genetic combination.

An insight into the magnitude of variability present in crop species is of utmost importance, as it provides the basis for effective selection. Genotypic variation is the component of variation that is due to the genotypic differences among individuals within a population or among populations within a species, and is the main concern of a plant breeder. Genetic variation may be measured in several ways. With univariate analyses, each variable is analyzed separately allowing for substantial overlapping of results to occur. Univariate statistical techniques such as analysis of variance do not explain how accessions differ when all measured variables are considered jointly. In Canonical Discriminant Analysis (CDA), a multivariate statistical technique, all independent variables are considered simultaneously in the differentiation of cultivars. This approach results in a more powerful comparison of populations than can be achieved with univariate analysis, provided the variables are correlated. Canonical discriminant analysis can separate among-population effects from within population effects by maximizing discrimination among populations when tested against the variation within populations (Riggs, 1973; Tai, 1989). Canonical variate analysis is discussed by Anderson (1958) has been frequently used in discriminating populations belonging to diverse genetic and geographic origin. It also serves as a pictorial presentation of the configuration of various groups. The information obtained from CDA can then additionally be used to group the accessions-populations into the smaller subgroups that are more similar to each other (Khattree and Naik, 2000). This CDA have been used in

**Author for Correspondence: E-mail:* bhavana.s.pandey1980@gmail.com

the assessment of genetic variation in tall fescue (*Festuca arundinacea* Schreb) (Vaylay and Santen, 2002) and hairy vetch (*Vicia villosa* Roth) (Yeater *et al.,* 2004).

Our objectives were to use the canonical discriminant analysis to study the genetic variation and selection of parents for further hybridization.

Materials and Methods

The experimental material consisted of 62 Indian varieties of soybean (Table 2). They were planted in a single row plot of three meter length with spacing of 60 × 10 cm in a Completely Randomized Block Design with two replications, at the Crop Research Centre of GB Pant University of Agriculture and Technology, Pantnagar, Uttarakhand. The observations were recorded on five randomly selected competitive plants. The agronomical characters were days to initiation of flowering, days to 50 per cent flowering, days to maturity, flowering span, reproductive phase, basal node height (cm), basal pod height (cm), plant height (cm), number of primary branches/plant, number of nodes/plant, number of pods /plant, number of seeds/pod, number of seeds /plant, grain yield/plant (g), 100-seed weight (g). In addition to agronomic characters two quality characters were also included in this study *viz.,* oil content (%) and protein content (%). Observations were subjected to CDA as discussed by Andersen (1958).The goal of this analysis is to arrive at cluster of accessions that display small within-cluster variation relative to the between-cluster variation.

Results and Discussion

Canonical discrimination is obtained by the ratio of variance among accessions to the variance within accessions (Rencher, 1992).

The extent of diversification, expressed as percentage contribution of first four canonical variate (Table 1) is the linear combination of the independent variables (characters) and is orthogonal to the other. A critical perusal of this table indicated that 34.40% of the total genetic diversity present in the genotypes was accounted by the first four canonical variates. All the four vectors contributed almost equally to total variation. Sum of all other variates contributed 65.60% of the total genetic variation. Therefore, the pre dominant axis of variation essentially remains in the sum of all other variates, playing a considerable role in further divergence. Other canonical variates also played a major part in determining total genetic diversity present in 62 varieties of soybean. This result is in agreement with the report of Shukla (1996). This may be due to the inclusion of much more diverse varieties and characters in this study.

The coefficients attached to the different characters in the first four canonical roots (vectors) are also shown in Table 1. The maximum importance of number of pods/plant followed by days to initiation of flowering and number of primary branches/plant in the primary differentiation between populations is reflected in the corresponding coefficients of the first canonical vectors. Similarly, 100-seed weight, days to initiation of flowering, number of pods/plant, basal pod height and grain yield /plant are reflected in their respective coefficients in the second canonical vectors indicating their relative importance in the secondary differentiation. In the third canonical vectors, protein percentage was of maximum importance followed by number of pods/plant and reproductive phase. Grain yield/plant, reproductive phase, number of primary branches/plant, 100-seed weight and plant height are reflected in their corresponding coefficients in the fourth canonical vectors indicating their relative importance in the further differentiations.

Table 1. Coefficients of first four canonical vectors in different soybean varieties

Character	Canonical vector			
	1	2	3	4
Days to initiation of flowering	0.458	-0.607	0.042	-0.061
Days to 50% flowering	0.175	-0.132	0.096	0.013
Days to maturity	0.113	-0.018	0.205	0.120
Flowering span	0.141	-0.101	0.251	-0.115
Reproductive phase	0.120	0.052	0.392	0.293
Basal node height	0.023	-0.267	0.027	0.082
Basal pod height	0.038	-0.349	-0.346	0.013
Plant height	0.302	-0.110	0.147	0.202
Number of primary branches/plant	0.393	0.156	0.247	0.278
Number of nodes/plant	0.044	0.145	-0.081	0.036
Number of pods/plant	0.566	0.352	-0.427	-0.223
Number of seeds/pod	0.280	0.089	-0.035	0.098
Number of seeds/plant	0.214	0.138	-0.331	-0.040
Grain yield/plant	0.023	0.340	-0.007	0.492
100-seed weight,	-0.067	0.936	-0.040	0.228
Oil content	0.046	-0.025	0.055	0.065
Protein content	0.109	0.273	0.472	-0.631
λ	60.080	58.830	57.230	55.670
%	8.920	8.730	8.490	8.560

Number of pods/plant appeared to be the most important in creating differentiation based on canonical variates. This was in line with the results obtained earlier in soybean by Das *et al.,* (2000); Ganesamurthy and Seshadari (2002) and Kayande and Patil (2009).

The mean value of the varieties for first two canonical roots (Z_1 and Z_2) is given in Table 2. The relative deposition of varieties in two dimensional space in Z_1 (Y-axis) and Z_2 (X-axis) (Fig. 1) shows the genetic divergence among the 62 varieties of soybean. The relative distribution of varieties reflected the broad genetic diversity among the genotypes. Based on the graph, 62 varieties were grouped into eighteen clusters (Table 3). Cluster I comprised on fifteen varieties while the Cluster II comprised of nine varieties. Cluster III and IV had eight and six varieties respectively. Cluster V had five varieties while Cluster VI, VII and VIII comprised three varieties each. Cluster XV had two varieties and remaining ten clusters were mono-genotypic. It was observed that the genotypes of different geographical origin were grouped together and the genotypes with same origin were included in different clusters. Since Cluster I comprised of varieties from nine different geographical areas. On the contrary, the varieties belonging to the same

Table 2. Mean value of the varieties for first 2 canonical roots (Z_1 and Z_2)

S. No.	Varieties/ genotypes	Developed/ released	Value		S. No.	Varieties/ genotypes	Developed/ released	Value	
			Z_1	Z_2				Z_1	Z_2
1	Alankar	Pantnagar	12.72	6.91	32	MACS 58	Pune	10.86	4.67
2	Ankur	Pantnagar	12.13	4.83	33	MAUS 2	Prabhani	10.59	4.09
3	Bragg	Pantnagar	10.91	5.24	34	MAUS 32	Prabhani	10.9	3.96
4	Birsa Soya-1	Ranchi	10.09	4.44	35	MAUS 47	Prabhani	11.22	5.59
5	Co 1	Coimbatore	10.69	4.21	36	Monetta	Prabhani	10.45	4.98
6	Co 2	Coimbatore	11.37	4.9	37	NRC 2	Indore	11.33	4.8
7	GS 1	Gujarat	11.56	4.23	38	NRC 12	Indore	11.48	5.94
8	Hardee	Karnataka	10.12	4.59	39	NRC 37	Indore	10.82	2.69
9	Himso 1563	Palampur	11.04	5.82	40	NRC 7	Indore	11.69	6.44
10	Improved Pelican	Karnataka	11.37	3.2	41	PS 1024	Pantnagar	11.19	5.61
11	Indira Soya 9	Raipur	10.36	3.96	42	PK 1029	Pantnagar	12.05	5.93
12	JS 2	Jabalpur	9.71	5.4	43	PK 262	Pantnagar	10.53	4.83
13	JS 335	Jabalpur	11.35	5.41	44	PK 308	Pantnagar	11.42	4.36
14	JS 71-05	Jabalpur	11.07	7.43	45	PK 327	Pantnagar	11.7	4.77
15	JS 72-280	Jabalpur	10.79	5.37	46	PK 416	Pantnagar	11.82	5.37
16	JS 72-44	Jabalpur	10.76	4.01	47	PK 471	Pantnagar	12.23	6.51
17	JS 75-46	Jabalpur	10.28	4.26	48	PK 472	Pantnagar	13.47	5.91
18	JS 76-205	Jabalpur	10.29	5.11	49	PK 564	Pantnagar	11.93	5.96
19	JS 79-81	Jabalpur	10.74	4.09	50	Punjab 1	Ludhiana	12.04	3.57
20	JS 80-21	Jabalpur	11.39	3.29	51	PUSA 16	New Delhi	10.39	4.23
21	JS 90-41	Jabalpur	9.69	5.06	52	PUSA 20	New Delhi	12.48	4.99
22	Kalitur	Bangalore	11.78	4.08	53	PUSA 22	New Delhi	11.5	4.73
23	KB 79	Bangalore	9.4	5.3	54	PUSA 24	New Delhi	10.32	4.73
24	KHSb 2	Karnataka	9.03	5.62	55	PUSA 37	New Delhi	11.2	4.2
25	Lee	Pune	11.28	3.99	56	PUSA 40	New Delhi	12.73	4.87
26	MACS 124	Pune	10.83	4.27	57	Samrat	Jabalpur	9.28	5.15
27	MACS 13	Pune	11.2	4.95	58	Shilajeet	Pantnagar	11.65	5.76
28	MACS 330	Pune	7.51	7.85	59	Shivalik	Pantnagar	11.25	4.72
29	MACS 450	Pune	10.71	4.54	60	SL 295	Ludhiana	12.26	6.56
30	MACS 57	Pune	12.35	4.31	61	T 49	Kanpur	11.42	3.56
31	MACS 754	Pune	10.23	4.25	62	VLS 47	Almora	12.3	5.14

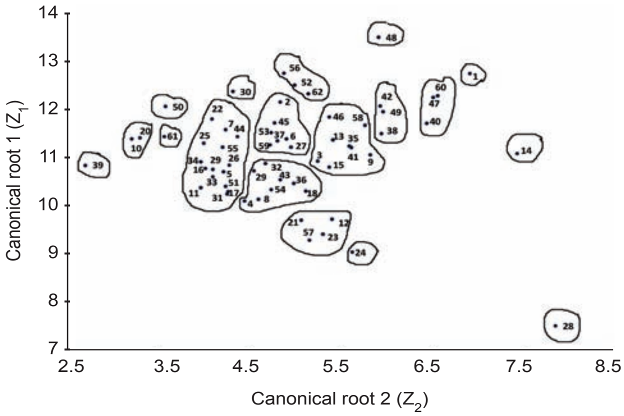

Fig. 1. Two dimensional presentation of divergence of genotypes of soybean using the first two canonical vectors as co-ordinate

geographical area of location fell into different clusters *viz.,* varieties developed at Pantnagar have been fallen into nine different clusters. The clustering pattern of the accessions showed that geographical diversity was not related with genetic diversity. Sharma (2000), Gawande *et al.,* (2002), Tyagi and Sethi (2011) and Patil *et al.,* (2011) reported that there exists no association between genetic diversity and geographical diversity.

The variety PK 472 forming a separate Cluster (IX) had maximum mean values for number of primary branches/plant (5.60), number of pods/ (101.30) and oil percentage (22.86). Alankar variety also had maximum mean values for number of seed/plant (159.90) and grain yield/plant (18.92g). Likewise JS 71-05 had maximum mean value for protein percentage (43.23). Thus, variety PK 472, Alankar and JS 71-05 holds great promise as a parent to obtain promising hybrids and create further variability for these characters.

Based on the genetic divergence analysis and cluster mean value (Table 3), it would be possible to point out some potential combinations, subject to the condition that environment maintain the relative expression of characters with regard to the genotypes. The potential combinations based on the canonical analysis were found to be NRC 37 × MACS 330, NRC 37 × JS 71-05, PK 472 × MACS 330, PK 472 × NRC 39, PK 472 × KHSb 2, T 49 × PK 472, T 49 × PK 471, T 49 × MACS 330, Samart × T 49, Samart × PK 472, NRC 37× Alankar and Alankar × T 49. These combinations should result in maximum hybrid vigour and highest number of useful segregants during the process of selection in the genotypes of soybean.

Table 3. Clustering pattern of cultivars on the basis of genetic divergence

Cluster group	Varieties	Number
I	GS -1, JS 72-44, Indira Soya 9, PK 308, JS 79-81, PUSA 16, Lee, Co 1, JS 75-46, PUSA 37, MAUS 2, MACS 754, MAUS 32, MACS 124, Kalitur	15
II	Ankur, Co 2, MACS 13, MACS 450, MACS 58, NRC 2, PK 327, PUSA 22, Shivalik	9
III	Bragg, Himso 1563, JS 335, JS 72-280, MAUS 47, PS 1024, PK 416, Shilajeet	8
IV	Birsa Soya-1, Hardee, JS 76-205, Monetta, PK 262, PUSA 24	6
V	JS 2, JS 90-41, KB 79, Samrat	4
VI	NRC 12, PK 1029, PK 564	3
VII	PUSA 20, PUSA 40, VLS 47	3
VIII	NRC 7, PK 471, SL 295	3
IX	PK 472	1
X	Alankar	1
XI	JS 71-05	1
XII	MACS 330	1
XIII	KHSb 2	1
XIV	NRC 37	1
XV	Improved Pelican, JS 80-21	2
XVI	T 49	1
XVII	Punjab 1	1
XVIII	MACS 57	1
TOTAL		62

It is concluded that significant genetic diversity was observed among the cultivars and, canonical discriminant analysis could identify the genetic variation and the most influential traits affecting genetic variations among the genotypes. Numbers of pods/plant and 100-seed weight were identified as the most effective traits in creating diversity amongst cultivars.

Acknowledgment

The authors thankfully acknowledge the Directorate of Research, GB Pant University of Agriculture and Technology, Pantnagar, Uttarakhand for the encouragement and help rendered during the course of this investigation.

Table 4. Cluster mean values value of the varieties for different characters

S. No.	Cluster group	Characters																
		Days to flowering initiation	Days to 50% Flowering	Days to Maturity	Flowering span	Repro-ductive phase	Basal node height (cm)	Basal pod height (cm)	Plant height (cm)	No. of primary branches / plant	No of nodes / plant	No of pods / plant	No of seeds / pod	No. of seeds / plant	Grain yield / plant (g)	100 seed weight (g)	Oil content (%)	Protein content (%)
1	I	52.90	57.67	123.50	11.77	68.80	1.71	13.79	69.43	3.59	14.60	42.75	1.77	66.17	5.88	9.35	20.30	37.02
2	II	49.28	56.72	123.39	10.94	71.78	1.66	13.91	71.39	3.59	15.76	54.40	1.87	88.77	8.84	10.23	20.67	39.59
3	III	49.69	53.56	122.00	10.63	73.81	1.57	14.11	65.83	3.95	15.01	50.78	1.93	85.93	9.80	11.62	20.15	40.14
4	IV	50.17	53.83	123.08	10.75	74.58	1.71	12.77	65.78	2.53	13.82	32.82	1.89	49.28	5.69	11.95	20.70	39.48
5	V	46.50	46.75	121.25	10.00	78.63	1.42	12.59	67.38	2.43	15.13	17.78	1.57	24.80	2.80	10.68	20.37	40.17
6	VI	53.50	52.17	122.50	11.00	75.50	1.75	12.32	73.37	3.97	14.20	62.20	1.90	104.37	13.80	13.55	20.79	40.05
7	VII	41.75	55.67	125.50	11.17	75.00	1.57	15.95	81.67	3.50	15.83	84.00	1.98	139.63	14.35	10.88	20.93	39.26
8	VIII	52.50	52.33	122.17	11.17	74.83	1.63	11.25	73.17	3.53	14.73	66.20	1.90	121.73	17.62	14.43	21.24	39.54
9	IX	49.50	58.00	124.00	15.00	74.5	1.66	10.30	70.30	5.60	15.20	101.30	1.88	105.90	17.25	11.62	22.86	39.2
10	X	55.50	59.50	120.50	10.00	65.00	1.23	7.85	53.00	4.80	15.10	83.60	2.00	159.90	18.92	11.89	20.79	39.55
11	XI	45.00	50.00	121.00	11.00	76.00	1.28	3.55	42.60	4.00	10.00	52.70	1.39	61.20	6.68	10.37	21.41	43.23
12	XII	21.00	25.00	79.00	9.00	58.00	1.00	5.85	34.00	1.40	10.50	1.40	1.00	1.40	0.10	10.17	18.62	40.69
13	XIII	44.00	48.50	122.00	10.00	78.00	1.80	8.60	50.00	2.00	10.60	5.00	1.42	7.90	1.00	13.75	20.75	40.43
14	XIV	53.50	59.00	122.50	12.50	69.00	2.60	16.60	74.70	2.90	15.30	41.60	1.89	53.60	4.77	9.17	21.54	37.28
15	XV	60.25	65.00	125.25	11.75	65.00	1.75	17..97	81.00	3.10	15.55	43.60	2.08	86.00	7.84	9.72	20.56	38.63
16	XVI	61.50	67.00	124.00	12.50	62.50	1.75	9.85	89.90	2.80	12.30	41.80	1.88	79.40	6.14	7.78	17.63	40.89
17	XVII	58.00	65.50	121.50	16.00	63.50	1.93	15.15	87.40	4.90	17.10	48.20	1.90	90.00	9.98	9.74	20.45	39.02
18	XVIII	57.00	60.50	119.50	11.00	62.50	1.80	1.7.3	73.30	5.20	19.30	80..90	1.87	138.20	12.16	8.95	20.36	38.51

References

Anderson TW (1958) An introduction to multivariate statistical analysis. Wiley, New York, USA, 374p.

Chung G and RJ Singh (2008) Broadening the genetic base of soybean: a multidisciplinary approach. *Critical Rev. Plant Sci.* **27:** 295-345.

Das SR, PN Harer and AB Biradar (2000) Genetic divergence and selection of genotypes in soybean. *J. Maha. Agril. Uni.* **25:** 250-252.

Ganesamurthy K and P Seshadari (2002) Genetic divergence studies in soybean (*Glycine max* L. Meriill). *Madaras Agric. J.* **89:** 18-21.

Gawande PP, S Patil, M Kumari and VJ Bailkiwad (2002) Selection of superior germplasm lines of soybean [Glycine max (L.) Merrill] for breeding. *J. Soils Crops* **12:** 236-240.

Kayande NV and SP Patil (2009) Genetic divergence in soybean (*Glycine max* L. Meriill). *Inter. J. Pl. Sci.* **4(1):** 218-228.

Khattree R and DN Naik (2000) Multivariate Data Reduction and Discrimination with SAS software, *SAS Institute Inc.,* Cary, NC, USA. 558p.

Patil SS, MR Naik, AB Patil and UR Ghodke (2011) Genetic diversity in soybean. *Legume Res.* **34:** 68-70.

Pushpendra, K Singh, BV Singh and MK Gupta (2008) Pant Soybean 1225 - An improved variety of soybean with broad genetic base. *Soybean Res.* **6:** 72-76.

Rencher AC (1992) Interpretation of canonical discriminant functions, canonical variants and principal components. *Am. Stat.* **46:** 217-225.

Riggs TJ (1973) The use of canonical analysis for selection within a cultivar of spring barley. *Ann. Appl. Biol.* **74:** 249-258.

Sharma SR (2000) Genetic divergence in soybean. *J. Oilseed Res.* **17:** 17-19.

Shukla S (1996) Studies on Genetic Diversity in the Indigenous and Exotic Germplasm Line of Soybean (*Glycine max* (L.) Merrill). *Thesis, M.Sc. GBPUA&T,* Pantnagar. 131p.

Tai GCC (1989) A proposal to improve the efficiency of index selection by 'rounding'. *Theor. Appl. Genet.* **78:** 798-800.

Tyagi SD and J Sethi (2011) Genetic Diversity Pattern in Soybean [*Glycine max* (L.) Merrill]. *Res. J. Agric. Sci.* **2:** 288-290.

Vaylay R and EV Santen (2002) Application of canonical discriminant analysis for the assessment of genetic variation in tall fescue. *Crop Sci.* **42:** 534-539.

Yeater KM, GA Bollero, DG Bullock, AL Rayburn and S Rodriguez-Z (2004) Assessment of genetic variation in hairy vetch using canonical discriminant analysis. *Crop Sci.* **44:** 185-189.