Electronic Journal of Plant Breeding

Research Article

Tailoring genetic diversity of greengram genotypes through principal component and cluster analysis

K. V. Patel*, D. J. Parmar, V. B. Kundaria, H. P. Patel and B. N. Patel

Pulse Research Station, Anand Agricultural University, Vadodara-389003, Gujarat, India
Department of Agricultural Statistics, B.A. College of Agriculture, AAU, Anand-388110, India
*E-Mail: kvpatel@aau.in

Abstract
One hundred genotypes of green gram were evaluated at the Pulse Research Station, AAU, Vadodara, Gujarat during summer 2019 for genetic diversity study based on six biometrical traits and to identify the best performing genotype suited to Gujarat. To determine the extent of variability existing among 100 green gram genotypes, Principal Component Analysis (PCA), cluster analysis and the inter-relationship existing between biometrical traits through Pearson's correlation analysis were carried out. The highest correlation was observed among the seeds per pod and pod length, pods per plant and the number of branches per plant. 100 seed weight had a negative and significant correlation with plant height, the number of branches per plant and pods per plant. According to the principal component analysis, two principal components (PCs) had eigenvalues more than unity and accounted to explain 67.37 per cent of the total variance that exists among six traits. PC1 and PC2 accounted for 39.88 and 27.49 per cent proportion of total variance, respectively. All traits were highly loading on PC1 except pod length and 100 seed weight which were loading on PC2. Three different groups of biometrical traits were observed in the component pattern. A hundred genotypes were classified into eleven clusters through the hierarchical cluster analysis method. Cluster I comprised a maximum number of genotypes followed by cluster VI, cluster II and cluster VII. Based on the cluster analysis, it could be recommended that crosses could be made between the genotypes of Cluster VIII and IX, Cluster VIII and X and Cluster X and XI which are distantly related. Lower plant height was observed in the genotypes under Cluster VIII and clusters X consisted of tall genotypes with a higher number of branches per plant. The highest pod length, pods per plant and seeds per pod were recorded under cluster IX. Cluster IV had a significantly higher 100 seed weight. Intraclass correlation (R^2) was the highest for pod length followed by plant height and seeds per pod indicating that these traits played an important role in divergence.

Key words
Green gram, biometrical traits, correlation, PCA, cluster analysis

INTRODUCTION

Vigna radiata (L.) Wilczek is an extensively cultivated pulse crop next to chickpea and pigeon pea. It is the most widely distributed and cultivated species among *Vigna* species. In tropical countries of the world, it is a pre-dominant source of protein for vegetarian people. Green gram is a short duration crop well suited for rotation and mixed farming, being a low water requirement crop suitable for drought tolerant crop, well adapted to a wide range of soil and also improve soil physical properties and fertility due to the presence of root nodules. Among pulse crops, the green gram is having average productivity of 425 kg/ha in Gujarat as well as in India which is quite low as compared to the other states of India and other countries . The major reason for the low productivity of pulses as well as green gram is poor agronomic management practices applied, inherent low yield potential of cultivar and susceptibility to viral disease.

For a successful crop improvement programme, exploitation of available genetic diversity is an important

pre-requisite for any plant breeding programme. Genetically diverse parents with the purpose of combining desirable recombinant for specific trait improvement followed by appropriate selection in segregating generation in a self-pollinated crop like green gram would result in the development of better cultivars. Keeping this point in view, the present study was undertaken to evaluate the available green gram germplasm using biometrical traits like plant height, the number of pods per plant, pod length, seeds per pod, the number of branches per plant and 100 seed weight through correlation analysis, principal component and cluster analysis for selection of diverse high yielding stable parent for attempting hybridization.

MATERIAL AND METHODS
One hundred genotypes of green gram were evaluated at the Pulse Research Station, AAU, Vadodara during summer 2019. The experiment was laid out in the augmented design. Each genotype has two row was of 4 m with the plant to plant spacing of 10 cm and row to row spacing of 45 cm. Six biometrical traits *viz.*, Plant height, Pod length, Seeds per pod, Branches per plant, Pods per plant and 100 seed weight that were recoded from five randomly selected plants and average value was worked out and used for further analysis.

PCA is a multivariate statistical method to summarize variation in multivariate samples with the construction of fewer artificial variables than the original set with minimal loss of information. Standardized values of six biometrical traits were used for PCA to find out the relative importance of different traits in capturing the genetic variation in the genotypes. The scree plot and eigenvalue were used to retain the number of a principal component which explain most of the variability present in the data. PCA 1 and PCA 2 scores of each genotype was taken to draw a scatter diagram and the genotypes scattering close to each other are taken to form a group or cluster. Cluster analysis of 100 genotypes based on biometrical traits was done using major principal components. The Euclidean distance was calculated using PROC DISTANCE and the ward clustering method was employed for clustering genotypes

RESULTS AND DISCUSSION
In any plant breeding programme, it is a pre-requisite to know the extent of the relationship among various biometrical traits to select plants for desirable characteristics. The knowledge of the inter relationship among yield components has been successfully exploited in crop improvement. Pearson (1901) correlation coefficient method was used to study the inter relationship among traits (**Table 1**). The highest significant positive correlation was observed between pod length and seeds per pod (0.582) followed by pods per plant and branches per plant. Plant height had a significant correlation with all traits except pod length. 100 seed weight had a negative and significant correlation with plant height, branches and pods per plant, whereas it showed a positive and significant correlation with pod length. The negligible and non-significant correlation was observed among pod length, plant height and branches per plant (Parameshwarappa, 2005; Parameshwarappa and Salimath,2007).

Table 1. Correlation among the different biometrical traits of greengram genotypes.

| Character	PH	PL	SP	BP	PP	TW
			Correlation Matrix			
Plant height (PH)	1.000					
Pod length (PL)	0.045	1.000				
Seeds per pod (SP)	0.410**	0.582**	1.000			
Branches per plant (BP)	0.451**	0.076	0.285**	1.000		
Pod per plant (PP)	0.469**	0.127	0.368**	0.473**	1.000	
100 seed weight (TW)	-0.218*	0.422**	0.010	-0.269**	-0.221*	1.000

*,** indicating significant @ 5% and 1% level, respectively

The PCA is highly effective as a data reduction tool. The principal component analysis (Multivariate analysis) was performed using six biometrical traits. In order to effectively utilizing the genotypes, characterization through principal component analysis provide a good screening of the different genotypes. The green gram genotypes showed significant variation for all the biometrical traits. The value of eigenvalue and variations explained by each of the principal components are presented in **Fig.1 and**

Table 2. The two principal components (PCs) had eigenvalues more than unity and accounted a total of 67.37 per cent of the total variance. Based on the scree plot, the first three PCs, PC1 accounted for high proportions of the total variance (39.88 %) followed by PC2 and PC3 which recorded 27.49 and 10.30 per cent of the total variation, respectively among genotypes. These results were in agreement with the findings of Divyaramakrishnan and Savithramma, (2014) and Kumar *et al*. (2018).

Table 2. Eigenvalues and percentage of variation in respect of six biometrical traits in greengram genotypes

Eigenvalues of the Correlation Matrix				
PC's	Eigenvalue	Difference	Proportion	Cumulative
1	2.3928	0.7434	0.3988	0.3988
2	1.6493	1.0312	0.2749	0.6737
3	0.6182	0.0507	0.1030	0.7767
4	0.5675	0.0488	0.0946	0.8713
5	0.5187	0.2652	0.0865	0.9578
6	0.2535		0.0422	1.0000
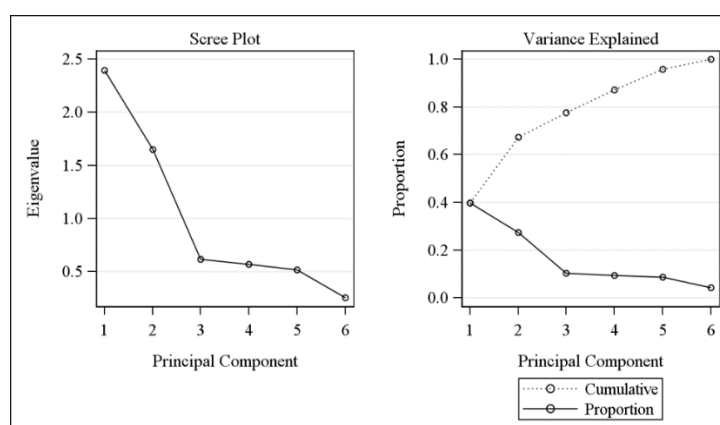
Fig.1. Scree plot for principal component analysis

Eigenvectors of the first two PCs were presented in **Table 3**. The results showed that the pod per plant had the highest positive value (0.495) followed by plant height (0.489), branches per plant and seeds per pod. In the second PC, pod length and 100 seed weight showed Eigenvector value of 0.661 and 0.601. (Basnet *et al.,* 2014; Manivannan *et al.,* 2015; Kumar *et al.,* 2018).

Table 3. Eigenvector values for PCs for various characters of greengram genotypes

Eigenvectors						
Characters/PCs	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
Plant height	**0.489**	-0.142	0.032	0.782	0.230	-0.274
Pod length	0.228	**0.661**	-0.100	-0.260	0.038	-0.657
Seeds per pod	**0.464**	0.363	-0.524	-0.006	0.057	0.612
Branches per plant	**0.467**	-0.193	0.474	-0.466	0.539	0.111
Pods per plant	**0.495**	-0.113	0.283	-0.101	-0.808	0.008
100 seed weight	-0.173	**0.601**	0.640	0.305	-0.001	0.326

The first two PCs were plotted against each other in biplot to observe the pattern and grouping of biometrical traits of green gram genotypes (**Fig. 2**). Plant height, pods per plant and branches per plant were observed in the same quadrant and they made one group whereas the second group of pod length and seeds per pod were placed in the second quadrant and 100 seed weight showed the third group and it appeared in the third quadrant (Kumar *et al.,* 2018). Hierarchical clustering using 100 green gram genotypes were grouped into eleven different clusters using a principal component score. The scores corresponding to six PCs were subjected to cluster analysis based on Euclidean distances and grouped by Ward method using PPROC CLUSTER (SAS 9.3). Cluster I had a maximum number of genotypes (25) followed by cluster VI (13), while clusters II and VII comprised   11-11 genotypes.   Cluster IV and VIII had 10 genotypes. Cluster III, V and IX had 7, 9 and 2 genotypes, respectively whereas cluster X and XI consist of one genotype in each cluster (**Table 4**). These results are in agreement with the findings of Sharma and Pawar (2007). Dhole and Reddy studied cluster analysis in mung bean. Seventeen mutants were grouped into six clusters. Diverse mutants can be used in mungbean improvement.
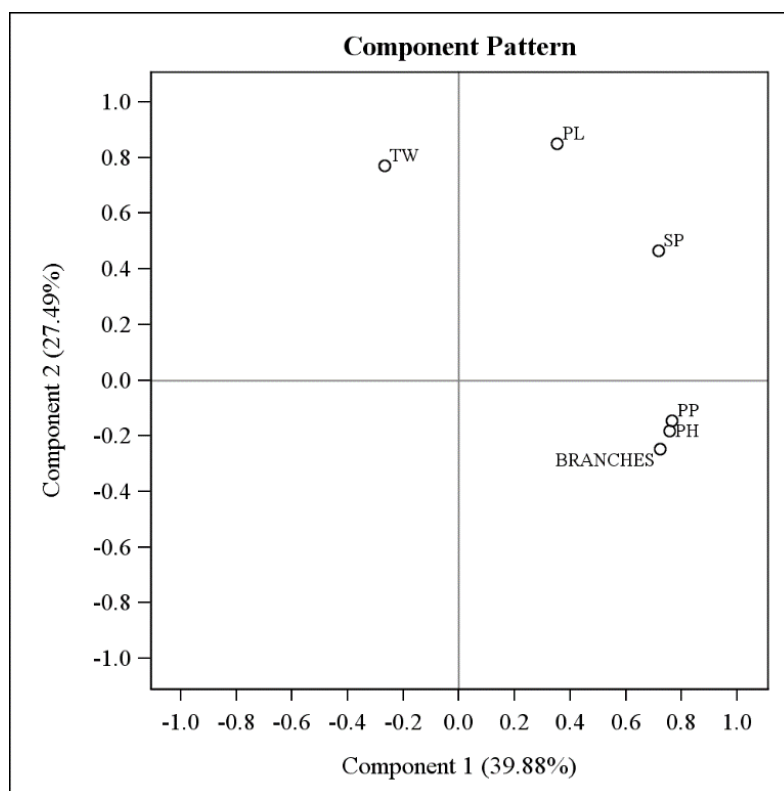
Fig. 2. Grouping of biometrical traits of green gram genotypes

Table 4. Distribution of green gram genotypes over different clusters based on six biometrical traits

Clusters	Number of genotypes	Name of Genotypes
I	25	VMG-6,VMG-12,VMG-13,VMG-18,VMG-21,VMG-25,VMG-27,VMG-28,VMG-31,VMG-34,VMG-40, VMG-41,VMG-48,VMG-50,VMG-56,VMG-60,VMG-63,VMG-65,VMG-71,VMG-77,VMG-89, VMG-90, VMG-96,VMG-97,VMG-98
II	11	VMG-37,VMG-45,VMG-46,VMG-58,VMG-107,VMG-108, VMG-111, VMG-112,VMG-115, VMG-116,VMG-117
III	7	VMG-8,VMG-51,VMG-72,VMG-84,VMG-87,VMG-88,VMG-100
IV	10	VMG-29,VMG-35,VMG-38,VMG-52,VMG-59,VMG-69,VMG-73,VMG-78,VMG-86,VMG-91
V	9	VMG-32,VMG-36,VMG-39,VMG-57,VMG-62,VMG-83,VMG-85,VMG-104,VMG-109
VI	13	VMG-3,VMG-5,VMG-22,VMG-30,VMG-33,VMG-55,VMG-68, VMG-79, VMG-80,VMG-81, VMG-82,VMG-94,VMG-113
VII	11	VMG-14,VMG-15,VMG-19,VMG-23,VMG-26,VMG-44,VMG-74,VMG-95,VMG-99, VMG-101,VMG-110
VIII	10	VMG-16,VMG-42,VMG-43,VMG-49,VMG-64,VMG-66,VMG-70,VMG-92,VMG-93,VMG-102
IX	2	VMG-75,VMG-76
X	1	VMG-103
XI	1	VMG-47

Based on the principal component I and II obtained from the PCA, a two dimensional scatter diagram (Prin1 – Prin2) was constructed using component score Prin1 as X axis and Prin2 as Y axis (**Fig. 3**). The position of the genotypes in the scatter diagram was apparently distributed into eleven groups, which indicated that considerable diversity exists among the genotypes. Convex of the hull occupied by the genotypes namely VMG 35 (cluster IV), VMG 93 (cluster VIII) VMG 47 (cluster XI), VMG 75,76 (cluster IX) and VMG 103 (Cluster X) showed the highest point

among the principal component. The average intra and inter-cluster distance values (Euclidean) were worked out and are presented in **Table 5**. The results showed that the most diverse clusters were VIII and IX (8.18) followed by Cluster X and XI (8.01), Cluster VIII and X (7.94). The genotypes related to diverse clusters can

be employed as distinct parents for the future breeding programme especially hybridization for the development of better segregants. The minimum inter-cluster value was observed between cluster III and V (2.16) which indicated that these groups were very less diverse from each other. (Divyaramakrishnan and Savithramma, 2014).
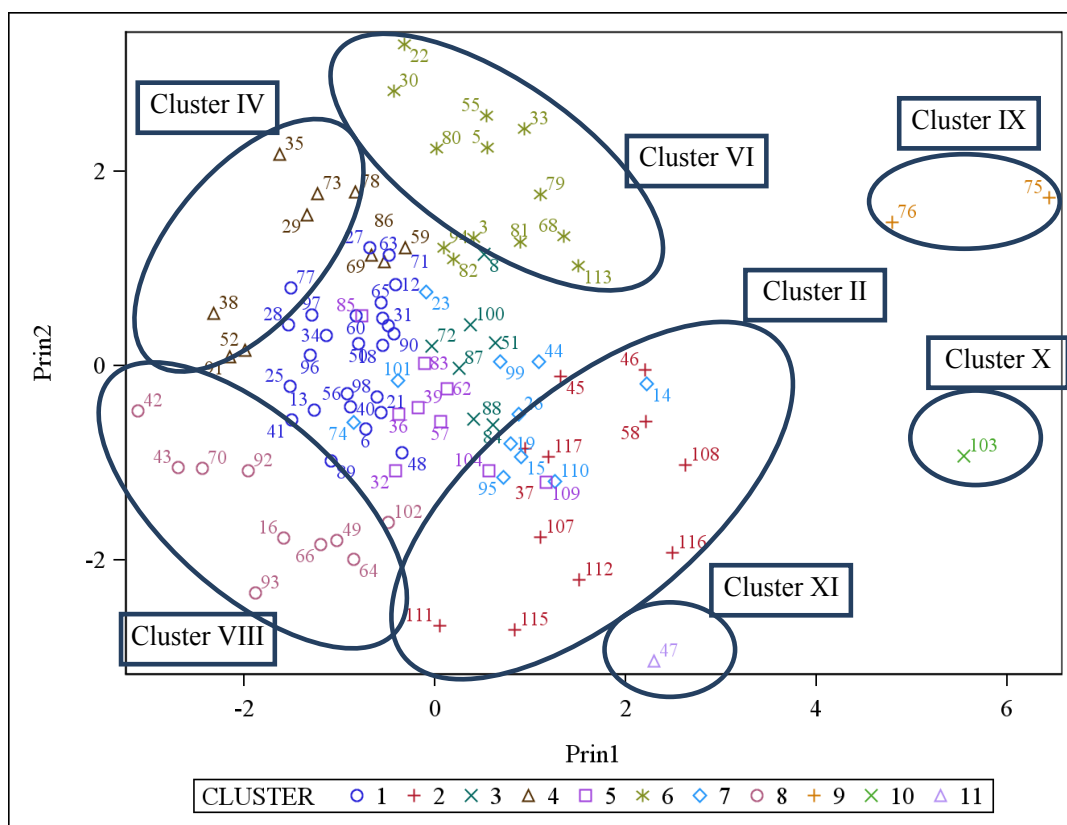

Fig. 3. Distribution of green gram genotypes based of two PCA score over different clusters

Table 5. Intra and inter cluster distance (Euclidean) for a hundred greengram genotype obtained by principal component analysis

Clusters	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
I	1.83	3.50	2.26	2.30	2.19	3.02	3.00	2.68	7.06	7.10	6.31
II		2.36	2.78	4.29	2.60	4.02	2.93	3.77	5.49	5.14	5.43
III			1.37	2.96	2.16	2.73	2.43	3.31	5.99	6.33	5.52
IV				1.83	2.62	2.87	3.45	3.27	7.18	7.81	7.49
V					1.51	3.17	2.53	2.59	6.20	6.40	6.08
VI						2.34	3.49	4.60	5.63	6.51	6.98
VII							2.07	3.64	6.18	6.10	6.24
VIII								1.98	8.18	7.94	6.45
IX									2.26	4.47	7.69
X										0.00	8.01
XI											0.00

The mean performance of all the characters in different clusters is presented in **Table 6**. The low plant height was observed in genotypes that fall under Cluster VIII whereas clusters X consisted of tall genotypes with the highest number of branches. The highest pod length, seeds per pod and pods per plant were recorded under cluster IX.

Cluster IV had a significantly higher 100 seed weight. Intra class correlation (R^2) was the highest for pod length followed by plant height and seeds per pod indicated that these traits play an important role in divergence (Pandiyan *et al.,* 2012; John *et al.*, 2015, Das and Baisakh, 2019).

Table 6. Different clusters means for sixbiometrical traits of greengram genotypes

Cluster	Plant height (cm)	Pod length (cm)	Seeds per pod	Branches per plant	Pods per plant	100 seed weight (g)
1	59.04	7.91	12.00	4.22	12.73	4.07
2	74.16	7.42	12.33	5.31	19.04	3.48
3	65.71	7.84	13.06	4.17	15.46	3.79
4	56.64	8.42	11.68	3.78	13.90	**4.79**
5	63.47	7.73	11.76	4.56	17.47	3.99
6	63.88	9.07	13.02	4.42	15.05	4.66
7	82.07	7.53	12.36	4.09	16.69	4.07
8	**53.82**	6.87	11.04	4.12	14.02	3.67
9	83.60	**9.88**	**14.50**	6.40	**28.10**	4.14
10	**104.20**	8.60	13.60	**8.00**	19.40	3.38
11	68.80	8.02	13.40	5.00	16.20	4.20
S.Em.	3.70	0.23	0.26	0.26	1.22	0.22
CD	10.41	0.64	0.74	0.73	3.44	0.61
CV%	10.5	5.3	4.0	10.8	14.6	9.8
R^2	0.85	0.88	0.85	0.82	0.81	0.73
CVb	24.80	14.04	9.43	23.13	29.76	16.34

The highest correlation was observed among the seeds per pod and pod length, pods per plant and branches per plant. 100 seed weight had a negative and significant correlation with plant height, branches per plant and pods per plant indicating that the development of bold seeded genotype with more branches per plant and pods per plant is very difficult. Results of the principal component analysis revealed that the first two principal components (PCs) had eigenvalues more than unity (1) and accounted for 67.37 per cent of the total variability describe among six traits. The genotypes were classified into eleven clusters through the hierarchical cluster analysis method. Cluster I composed a maximum number of genotypes followed by cluster VI, cluster II and cluster VII, which justifies that crosses could be made between the genotypes of Cluster VIII and IX, Cluster VIII and X and Cluster X and XI. Intraclass correlation (R^2) was the highest for pod length followed by plant height and seed per pod indicating that these traits played an important role in genetic divergence in green gram for seed yield improvement.

REFERENCES

Basnet, K.M., Adhikari, N.R. and Pandey, M.P. 2014. Multivariate analysis among the Napalese and exotic mungbean (*Vigna radiata* (L.) Wilczek) genotypes based on the qualitative parameters. *Univ. J. Agril. Res,*. **2(5)**:147-155.

Das Tapas Ranjan and Bhabendra Baisakh. (2019). Genetic divergence among mutant genotypes of greengram [Vigna radiata (L.) Wilczek] *Electronic Journal of Plant Breeding*, **10** (1): 160 – 16.

Dhole V. J. and K. S. Reddy (2018). Genetic analysis and variability studies in mutants induced through electron beam and gamma rays in mungbean (Vigna radiata L. Wilczek). *Electronic Journal of Plant Breeding*, **9** (1) : 304 – 312.

Divyaramakrishnan, C.K. and Savithramma D.L. 2014. Tailoring genetic diversity of mungbean [*Vigna radiata* (L.) Wilczek] germplasm through principal component and cluster analysis for yield and yield related traits. *Int. J. Agron. and Agril. Res.* **5 (2):** 94-102.

John Kingsly N.B., Packkiaraj, D., Pandiyan M. and Senthil. N. 2015. Tailoring genetic diversity of mungbean (*Vigna radiata* (L.) Wilczek) germplasm through cluster analysis for yield and yield related traits. *Trends in Biosciences* **8(12):** 3239-3244.

Kumar Umesh V. and Sarkar, K.K. 2018. Tailoring the genetic diversity of mung bean through cluster analysis for yield attributing traits to obtain efficient hybrids. *Int.J.Curr.Microbiol.App.Sci.,* **7(7)**: 1101-1108. [Cross Ref]

Manivannan A., Anandakumar, C. R., Ushakumari, R and Dahiya, G. S. 2015. Genetic diversity of guar genotypes (*Cyamopsis tetragonoloba* (L.) Taub.) based on agro-morphological traits *Bangl. J. Bot.* **44(1)**: 59-65. [Cross Ref]

Kumar Naveen Jakhar and Arjun Kumar 2018. Principal component analysis and character association for yield components in greengram [*Vigna radiata* (L.) Wilczek] genotypes. *J. Pharma. and Phytochem.* **7(2)**: 3665-3669.

Pandiyan M., Senthil N., Packiaraj D., Thanga Pandian R., Suresh R. and Jagadeesh S. 2012. Characterization and evaluation of 646 greengram (*Vigna radiata* L.) genotypes for constituting core collection. *Wudpecker J. Agri.l Res.,* **1(8):**294-301.

Parameshwarappa, S. G. and Salimath, P. M. 2007. Studies on genetic variability,character association and path coefficient analysis in greengram. *Crop Research-HISAR*, **34(1/3)**: 195.

Parameshwarappa S. G. 2005. Genetic variability character association and path coefficient analysis in green gram. *Karnataka Journal Agricultural Science*, **4**: 1090-1092.

Pearson K. 1901. On lines and planes of closest fit to systems of points in space. *Philos Mag. A.* **6**:559–572. [Cross Ref]

Sharma, V. and Pawar, I. S. 2007. Genetic divergence, for yield its components and quality traits in bread wheat. *Haryana Agric. Uni. J. Res.* **37:** 27-31.