Article AR2078 DOI: HTTPS://DOI.ORG/10.23910/IJBSM/2020.11.2.2078

International Journal of Bio-resource and Stress Management

April 2020

Print ISSN 0976-3988
Online ISSN 0976-4038

IJBSM 2020, 11(2):138-147

Research Article | *Natural Resource Management*

Genetic Diversity Study through K-means Clustering in Germplasm Accessions of Green gram [*Vigna radiata* (L.)] Under Drought Condition

Kanavi, M. S. P.[1*], Prakash Koler[2], Somu, G.[3] and N. Marappa[4]

[1]Dept.of Genetics and Plant Breeding, [2]Dept. Crop Physiology, College of Agriculture, Hassan, University of Agricultural Sciences, Bangalore, Karnataka (57 3225), India

[3]Assistant Breeder, Dept. of Genetics and Plant Breeding, AICRP on Sorghum, Chamarajanagara, University of Agricultural Sciences, Bangalore, Karnataka (571 313), India

[3]Dept. of Genetics and Plant Breeding, College of Agriculture, G.K.V.K., University of Agricultural Sciences, Bangalore, Karnataka (560 065), India

Open Access

Corresponding Author

Kanavi, M.S.P.

e-mail: kanavi.uasb@gmail.com

Citation: Kanavi et al., 2020. Genetic Diversity Study through K-means clustering in Germplasm Accessions of Green gram [*Vigna radiata* (L.)] Under Drought Condition. International Journal of Bio-resource and Stress Management 2020, 11(2), 138-147. HTTPS://DOI.ORG/10.23910/IJBSM/2020.11.2.2078.

Data Availability Statement: Legal restrictions are imposed on the public sharing of raw data. However, authors have full right to transfer or share the data in raw form upon request subject to either meeting the conditions of the original consents and the original research study. Further, access of data needs to meet whether the user complies with the ethical and legal obligations as data controllers to allow for secondary use of the data outside of the original study.

Funding: The research was conducted with the kind and supports from Institute

Conflict of interests: The authors have declared that no conflict of interest exists.

Acknowledgement: Kanavi, M. S. P., thanks Director of Research, University of Agricultural Sciences, Bangalore for giving financial assistance to carry out the research work

Abstract

An experiment was conducted to evaluate 205 green gram germplasm accessions along with five check entries for drought tolerance using augmented design during summer 2015 by imposing drought stress condition. Observations were recorded on 17 quantitative traits. ANOVA revealed high significant differences among germplasm accessions for yield, yield component traits and also for drought tolerance traits. Mean squares attributable to 'Genotypes *vs* check entries' were significant for all the traits except seeds per pod and relative water content. Based on K-means clustering, all the 205 germplasm accessions were grouped into seven different clusters. Cluster V was the largest with 38 genotypes followed by cluster I with 36, cluster III and VII with 28, cluster II with 27, cluster IV with 25 and cluster VI with 23 genotypes. The mode of distribution of genotypes coming from different geographical regions into various clusters was at random indicating that the genotypes originating from different agro-climatic regions grouped together into different clusters showing no parallelism between genetic diversity and geographical distribution. The maximum inter cluster distance was recorded between the clusters I and VI (208.17) followed by cluster V and VI (168.52). The minimum inter cluster distance was recorded between the clusters IV and V (45.01) followed by cluster IV and VII (46.97). The maximum intra cluster distance was recorded for the cluster VI (208.17) followed by cluster IV (160.40). The minimum intra cluster distance was recorded for the cluster IV (45.01) followed by cluster V (52.55).

Keywords: Green gram, drought tolerance, genetic diversity, k-means clustering

1. Introduction

Green gram [*Vigna radiata* (L.) Wilczek] also known as mung bean is an important short duration pulse crop of the tropical and subtropical countries of the World. Green gram is the third most important pulse crop of India after chickpea and red gram. It belongs to papilionoid subfamily of the Fabaceae family and has a diploid chromosome number of 2*n*=2*x*=22. The word "Pulse" is derived from the Latin word "Puls" meaning pottage i.e. seeds boiled to make porridge or thick soup (Singh

Article History

RECEIVED in 29th February 2020 RECEIVED in revised form 02nd April 2020 ACCEPTED in final form 15th April 2020

et al., 2018). The seeds of green gram are rich in minerals like phosphorus, calcium, vitamins and also contain higher levels of folate and iron than most other legumes (Keatinge et al., 2011). The protein content of pulses are twice that of cereals (20-25%) and almost equal to that of meat and poultry hence commonly pulses are called as the poor man's meat (Reddy, 2009). India being major pulse producing country in the World which shares 30-35% and 27-28% of the total area and production respectively. Average productivity of mung bean in India is one of the lowest compared to World average. The reason attributable to lower productivity of green gram in India is that the crop is mainly grown as a fallow crop in rabi or late rabi season utilizing available residual soil moisture after harvesting main kharif crop. Hence crop is expected to experience several kinds of droughts during its cropping period. Drought is the major constraint for green gram production due to insufficient and erratic rainfall in India.

Genetic diversity refers to the number of different alleles of all genes and the frequency with which they appear in the population. In green gram, the morphological characterization of accessions belonging to cultivated species reveal high genetic variability for a trait other than genetic diversity. Murthy and Arunachalam (1966) emphasized the importance of genetic diversity in selection of parents for hybridization programme in different crops. Genetic diversity present in the germplasm accessions is an important tool for any plant breeding program (Azam et al., 2018). The assessment of genetic variation would provide us a correct picture of the extent of genetic variation, further helping us to improve the genotypes responses to biotic and abiotic stresses (Panigrahi and Baisakh, 2014). The genetic variability offers a working bench for selection intensity and direction which is determined by the crop breeders according to breeding objectives for crop improvement activities in mungbean. Genetic diversity is one of the critical criteria for selection of parents in the hybridization program to isolate best genotypes from transgressive segregants. Cluster analysis in green gram would definitely help plant breeders to identify genetically diverse parents falling in different clusters (Umesh et al., 2017)

Clustering is a technique where millions of data points are grouped together to form a cluster. Cluster analysis or clustering is to group, categorize or classify a set of objects into many subsets, called clusters, in such a way that the items inside one subset are more "similar" to each other, while "dissimilar" to items inside other subsets. Therefore there must be a way to distinguish between "dissimilar" and "similar" items. K-means clustering is very important and basic clustering technique through which data points are analyzed. K-means is one of the most widely used algorithm for clustering with known sets of median points. Clustering can be used in an exploratory manner to discover meaningful groupings within a data set, or it can serve as the starting point for more advanced analysis (Wang et al., 2019). K-means clustering is an unsupervised machine learning algorithm. It is preferred as the attractiveness lies in its efficiency with O(n*K*i*a), where n, K, I and a equals number of data points, clusters, iterations and attributes respectively. Assessment of genetic diversity is must for any plant breeding programme to identify genetically diverse parents to be involved in hybridization programmes. K-means clustering is a very powerful technique to assess genetic diversity which creates genetically diverse clusters / heterotic groups based on genetic distances between germplasm accessions. Once the heterotic groups are created, then it is easy to identify clusters which are genetically very distant and the germplasm accessions falling in this clusters are also genetically very diverse. Thus it becomes easy for plant breeders to identify genetically diverse germplasm accessions which in-turn will serve as parental lines in crossing programme. This research was carried out with a purpose to identify genetically diverse drought tolerant genotypes which can be later used as parental lines in plant breeding programmes to develop drought tolerant genotypes.

2. Materials and Methods

The experiment was conducted at experimental plot of College of Agriculture, Hassan, University of Agricultural Sciences, Bengaluru, India. The experimental site is geographically located at Southern Transitional Zone (Zone-7) of Karnataka with an altitude of 827 m above Mean Sea Level (MSL) and at 33' N latitude and 75° 33' to 76° E38' longitude. The study material consisted of 205 germplasm accessions collected from different research institutions / organizations representing different agro-climatic zones. List of germplasm accessions used in the study with their source is given in Table 1.

2.1. Layout of the experiment

The experiment was conducted in an Augmented Randomized Complete Block Design with 205 germplasm accessions. As per the augmented RCBD, the check entries were replicated twice randomly in each block. There were 5 blocks, each block had 5 plots of size 3x3 m^2 thus each block size was 15 m^2. The gross area of experimental plot was 75 m^2. The row spacing was 30 cm and inter plant distance was 10 cm. The experiment was conducted during *summer* 2015. Recommended crop production practices were followed to raise healthy crop.

2.2. Imposing drought condition

Drought condition was imposed by withholding irrigation 25 days after sowing (Bangar et al., 2019). Since the experiment was conducted during *summer* season, there were no unpredicted rains during the entire cropping period hence the drought condition was effectively imposed. The rainfall data of experimental site during the cropping period is given in Table 2.

2.3. Plant sampling and data collection

Observations were recorded on five randomly chosen competitive plants from each germplasm accession for all

Table 1: List of germplasm accessions used in the study and their source

Sl. No.	Germplasm	Location	Sl. No.	Germplasm	Location	Sl. No.	Germplasm	Location
1	KM13-16	ARS, Bidar	41	AKL-103	NBPGR, Akola	81	PM-112	TNAU, Coimbatore
2	KM13-19	ARS, Bidar	42	AKL-39	NBPGR, Akola	82	LGG-578	NBPGR, Akola
3	KM13-39	ARS, Bidar	43	AKL-106	NBPGR, Akola	83	LGG-563	NBPGR, Akola
4	GG13-7	ARS, Bidar	44	AKL-225	NBPGR, Akola	84	LGG-594	NBPGR, Akola
5	GG13-6	ARS, Bidar	45	AKL-95	NBPGR, Akola	85	TM-96-2	NBPGR, Akola
6	KM13-44	ARS, Bidar	46	AKL-194	NBPGR, Akola	86	LGG-593	NBPGR, Akola
7	GG13-10	ARS, Bidar	47	AKL-212	NBPGR, Akola	87	LGG-591	NBPGR, Akola
8	SML-668	ARS, Bidar	48	AKL-195	NBPGR, Akola	88	PM-115	NBPGR, Akola
9	KM13-9	ARS, Bidar	49	AKL-211	NBPGR, Akola	89	LGG-587	NBPGR, Akola
10	IPM99-125	ARS, Bidar	50	KM13-11	ARS, Bidar	90	PM-113	NBPGR, Akola
11	LGG-596	RARS, Guntur	51	KM13-30	ARS, Bidar	91	LGG-586	NBPGR, Akola
12	LGG-572	RARS, Guntur	52	KM13-45	ARS, Bidar	92	IC-436775	NBPGR, Akola
13	LGG-450	RARS, Guntur	53	KM13-18	ARS, Bidar	93	IC-413311	NBPGR, Akola
14	LGG-583	RARS, Guntur	54	KM13-5	ARS, Bidar	94	IC-398984	NBPGR, Akola
15	LGG-590	RARS, Guntur	55	KM13-02	ARS, Bidar	95	IC-436767	NBPGR, Akola
16	LGG-588	RARS, Guntur	56	KM13-37	ARS, Bidar	96	IC-436573	NBPGR, Akola
17	LGG-589	RARS, Guntur	57	KM13-23	ARS, Bidar	97	LGG-584	NBPGR, Akola
18	LGG-579	RARS, Guntur	58	KM13-55	ARS, Bidar	98	LGG-592	NBPGR, Akola
19	LGG-562	RARS, Guntur	59	KM13-12	ARS, Bidar	99	LGG-555	NBPGR, Akola
20	LGG-582	RARS, Guntur	60	GG13-9	ARS, Bidar	100	LGG-564	NBPGR, Akola
21	LGG-585	RARS, Guntur	61	KM13-49	ARS, Bidar	101	LGG-460	RARS, Guntur
22	AKL-170	NBPGR, Akola	62	GG13-4	ARS, Bidar	102	LGG-595	RARS, Guntur
23	PLM-110	UAS, Bangalore	63	GG13-54	ARS, Bidar	103	LGG-566	RARS, Guntur
24	LGG-577	RARS, Guntur	64	KM13-20	ARS, Bidar	104	IC-553514	IIPR, Kanpur
25	IC-436624	IIPR, Kanpur	65	GG13-5	ARS, Bidar	105	IC-413319	IIPR, Kanpur
26	IC-436723	IIPR, Kanpur	66	Chinamung	ARS, Bidar	106	IC-436542	IIPR, Kanpur
27	IC-413316	IIPR, Kanpur	67	GG13-2	ARS, Bidar	107	IC-546493	IIPR, Kanpur
28	IC-436746	IIPR, Kanpur	68	KM13-26	ARS, Bidar	108	IC-436594	IIPR, Kanpur
29	VGG10-010	TNAU, Coimbatore	69	KM13-47	ARS, Bidar	109	IC-436630	IIPR, Kanpur
30	VGG04-011	TNAU, Coimbatore	70	KM13-41	ARS, Bidar	110	IC-436668	IIPR, Kanpur
31	VGG04-007	TNAU, Coimbatore	71	KM13-11	ARS, Bidar	111	IC-436555	IIPR, Kanpur
32	COGG-93	TNAU, Coimbatore	72	KM13-42	ARS, Bidar	112	IC-413314	IIPR, Kanpur
33	VBNGG-2	TNAU, Coimbatore	73	GG13-11	ARS, Bidar	113	AKL-20	NBPGR, Akola
34	TARM-2013	TNAU, Coimbatore	74	GG13-8	ARS, Bidar	114	AKL-89	NBPGR, Akola
35	VGG04-005	TNAU, Coimbatore	75	GG13-12	ARS, Bidar	115	AKL-228	NBPGR, Akola
36	COGG-920	TNAU, Coimbatore	76	KM13-48	ARS, Bidar	116	AKL-184	NBPGR, Akola
37	VGG07-003	TNAU, Coimbatore	77	IPM2-3	ARS, Bidar	117	AKL-182	NBPGR, Akola
38	VGG10-002	TNAU, Coimbatore	78	IPM2-14	ARS, Bidar	118	AKL-230	NBPGR, Akola
39	VGG-112	TNAU, Coimbatore	79	PDM-139	ARS, Bidar	119	AKL-229	NBPGR, Akola
40	IC-92048	NBPGR, Akola	80	LGG-580	RARS, Guntur	120	AKL-86	NBPGR, Akola

Sl. No.	Germplasm	Location	Sl. No.	Germplasm	Location	Sl. No.	Germplasm	Location
121	IC-436646	IIPR, Kanpur	155	AKL-189	NBPGR, Akola	175	AKL-105	NBPGR, Akola
122	IC-343964	IIPR, Kanpur	156	AKL-168	NBPGR, Akola	176	AKL-213	NBPGR, Akola
123	IC-436528	IIPR, Kanpur	157	AKL-218	NBPGR, Akola	177	AKL-169	NBPGR, Akola
124	IC-436723	IIPR, Kanpur	158	AKL-179	NBPGR, Akola	178	AKL-220	NBPGR, Akola
125	IC-546491	IIPR, Kanpur	159	AKL-185	NBPGR, Akola	179	AKL-84	NBPGR, Akola
126	IC-546481	IIPR, Kanpur	160	AKL-163	NBPGR, Akola	180	AKL-82	NBPGR, Akola
127	IC-398988	IIPR, Kanpur	161	COGG-912	TNAU, Coimbatore	181	AKL-97	NBPGR, Akola
128	VGG10-005	TNAU, Coimbatore	162	IC-73451	NBPGR, Akola	182	AKL-226	NBPGR, Akola
129	VBN-223	TNAU, Coimbatore	163	IC-105690	NBPGR, Akola	183	AKL-24	NBPGR, Akola
130	COGG-912	TNAU, Coimbatore	164	IC-73534	NBPGR, Akola	184	AKL-174	NBPGR, Akola
131	VBN(G9)-3	TNAU, Coimbatore	165	IC-73412	NBPGR, Akola	185	AKL-161	NBPGR, Akola
132	ML-1165	TNAU, Coimbatore	166	IC-39605	NBPGR, Akola	186	AKL-180	NBPGR, Akola
133	VGG04-025	TNAU, Coimbatore	167	IC-73472	NBPGR, Akola	187	AKL-222	NBPGR, Akola
134	VGG04-004	TNAU, Coimbatore	168	IC-92053	NBPGR, Akola	188	AKL-187	NBPGR, Akola
135	VGG04-149	TNAU, Coimbatore	169	IC-73779	NBPGR, Akola	189	AKL-216	NBPGR, Akola
136	COGG-954	TNAU, Coimbatore	170	IC-73462	NBPGR, Akola	190	AKL-29	NBPGR, Akola
137	VGG08-002	TNAU, Coimbatore	171	IC-118992	NBPGR, Akola	191	AKL-90	NBPGR, Akola
138	VBN-1	TNAU, Coimbatore	172	IC-53783	NBPGR, Akola	192	AKL-227	NBPGR, Akola
139	VGG-119	TNAU, Coimbatore	173	IC-73456	NBPGR, Akola	193	AKL-200	NBPGR, Akola
140	VC3890-A	TNAU, Coimbatore	174	IC-73458	NBPGR, Akola	194	AKL-92	NBPGR, Akola
141	DGGV-4	UAS, Raichur	175	AKL-105	NBPGR, Akola	195	AKL-183	NBPGR, Akola
142	KPS-1	UAS, Raichur	176	AKL-213	NBPGR, Akola	196	AKL-176	NBPGR, Akola
143	CGG-973	UAS, Raichur	177	AKL-169	NBPGR, Akola	197	AKL-191	NBPGR, Akola
144	CN9-5	UAS, Raichur	178	AKL-220	NBPGR, Akola	198	AKL-165	NBPGR, Akola
145	KPS-2	UAS, Raichur	179	AKL-84	NBPGR, Akola	199	AKL-164	NBPGR, Akola
146	VC-6173	UAS, Raichur	180	AKL-82	NBPGR, Akola	200	AKL-192	NBPGR, Akola
147	VC-6368	UAS, Raichur	181	AKL-97	NBPGR, Akola	201	BGS-9	Check entry
148	CO-6	UAS, Raichur	182	AKL-226	NBPGR, Akola	202	DGG-1	Check entry
149	Harsha	UAS, Raichur	183	AKL-24	NBPGR, Akola	203	PDM 84-178	Check entry
150	PLM-92	UAS, Bangalore	170	IC-73462	NBPGR, Akola			
151	MH-709	UAS, Raichur	171	IC-118992	NBPGR, Akola	204	PS-16	Check entry
152	LGG-460	RARS, Guntur	172	IC-53783	NBPGR, Akola	205	KKM-3	Check entry
153	KGS-5	UAS, Raichur	173	IC-73456	NBPGR, Akola			
154	Barimung-4	UAS, Raichur	174	IC-73458	NBPGR, Akola			

the characters except days to 50% flowering and days to maturity, which were recorded on plot basis. The values of five competitive plants were averaged and expressed as mean of the respective characters. The observations were taken on the traits like; Days to 50% flowering, Days to maturity, Plant height (cm), Clusters plant^{-1}, Pods cluster^{-1}, Pods per plant, Pod length (cm), Seeds pod^{-1}, test weight, Threshing %, Harvest index (%), SCMR (SPAD Chlorophyll meter reading), Leaf water potential (Mpa), Proline content (µg g^{-1}), Relative water content, Specific leaf area and Seed yield per plant.

2.4. Statistical analysis

2.4.1. Analysis of variance (ANOVA)

The quantitative trait mean value of five randomly selected plants in each of the genotype and check entries were used for statistical analysis. ANOVA was performed to partition the total variation among genotypes and check entries into sources attributable to 'Genotypes+Check entries',

Table 2: Meteorological data of experimental site for the year 2015

Months	Temperature (°C)			Relative humidity (%)	Rainfall (mm)
	Maximum	Minimum	Average		
January	28.25	15.00	21.32	61.03	0.59
February	30.35	15.25	23.10	50.72	Nil
March	31.70	19.50	25.34	58.70	2 mm (25.03.2015)
April	32.50	21.25	25.87	66.55	Nil

Genotypes', Check entries' and Genotypes vs check entries', following the augmented design as suggested by Federer (1956) using statistical package for augmented design SAS version 9.3 and IndoStat. The adjusted trait mean of each of the genotype was estimated (Federer, 1956) and the same was used for all subsequent statistical analysis.

2.4.2. K-means clustering

The germplasm accessions were classified following 'k-means clustering' model as explained by Macqueen (1967) and Forgy (1965). K-means cluster analysis was performed in SAS 9.3 version and NCSS statistical software. The trait means and variances were estimated in each cluster and tested for their homogeneity across the cluster using 'F' test and 'Levene' test (Levene, 1960)

The test statistic (W) for Leven's test was computed as,

$$W = \frac{(N-K)\sum_{i=1}^{K} Ni\,(Zi.-Z..)^2}{(K-1)\sum_{i=1}^{K}\sum_{j=1}^{Ni}(Zij-Zi.)^2}\ \bar{Y}ij-\bar{Y}i$$

Where,

$Zi. = \frac{1}{Ni}\sum_{j=1}^{Ni} Zij$ is the mean of Zij for ith cluster

$Zij = |\bar{Y}ij - \bar{Y}i|$, Yi is the mean of ith cluster

$Z.. = \frac{1}{Ni}\sum_{i=1}^{K}\sum_{j=1}^{Ni} Zij$ is the mean of Zij

3. Results and Discussion

3.1. Analysis of variance (ANOVA)

Analysis of variance revealed highly significant mean squares attributable to germplasm accessions for all the traits. Significant mean squares were recorded for all the traits. (Table 3). Mean squares attributable to 'Genotypes *vs* check entries' were significant for all the traits except seeds per pod and relative water content. These results suggest significant differences among the germplasm accessions. The germplasm accessions as group differed significantly for all of the traits under investigation, similarly, check entries as group differed significantly for most of the traits under study.

Table 3: Summary of augmented ANOVA for grain yield and component traits of germplasm accessions under drought condition

Sources of Variations	DF	DFF	DM	PH	CPP	PPC	PPP	PL	SPP	TW
Blocks (b)	4	14.74**	8.18***	65.31**	2.23**	0.11*	25.23**	1.49**	5.05**	1.77**
Entries (e) (Genotypes + Checks)	204	17.10**	18.01**	84.47**	3.60**	0.51**	72.94**	0.75**	2.70**	0.35**
Checks	4	34.57**	37.01**	22.56**	1.40**	0.42**	12.50**	0.87**	3.98**	0.81**
Genotypes	199	14.215**	15.14**	85.71**	3.67**	0.51**	73.91**	0.73**	2.69**	0.31**
Checks vs Genotypes	1	521.64**	513.06**	85.01**	0.16**	1.45**	121.60**	4.52**	0.03	5.42**
Error	16	1.32	0.74	0.98	0.04	0.02	0.98	0.009	0.05	0.05

Sources of Variations	DF	TP	HI	SCMR	LWP	PC	RWC	SLA	SYPP
Blocks (b)	4	37.12*	247.54**	396.55**	1.17**	470.90**	423.68*	4067.34*	2.11**
Entries (e) (Genotypes + Checks)	204	37.20**	54.41*	98.71**	2.45**	1707.90**	425.40**	4283.10**	7.01**
Checks	4	17.09	64.39*	24.49	0.82**	942.07**	63.06	1924.20	3.76**
Genotypes	199	27.67*	53.01*	79.58*	2.33**	1712.67**	433.68**	4294.15**	7.10**
Checks vs Genotypes	1	2014.79**	293.20**	4203.25**	32.57**	3822.09**	227.32	11518.68**	0.42*
Error	16	9.83	19.57	31.14	0.03	1.48	130.64	1339.95	0.09

*Significant at *p*=0.05, ** Significant at *p*=0.01; DFF : Days to 50% flowering; Pods plant^{-1}; HI : Harvest index (%); SLA : Specific leaf area; DM : Days to maturity; PL : Pod length (cm); SCMR : SPAD Chlorophyll meter reading; SYPP : Seed yield plant^{-1}; PH : Plant height (cm); SPP : Seeds per pod; LWP : Leaf water potential (Mpa); CPP : Cluster plant^{-1}; TW: test weight (g); PC : Proline content (μg g^{-1}); PPC : Pods cluster^{-1}; TP : Threshing %; RWC: Relative water content (%)

3.2. K-means clustering

K-means clustering intends to partition *n* objects into *k* clusters in which each object belongs to the cluster with nearest mean. K-means is a centroid based clustering algorithm. 'K' represents the number of clusters, and it is also an input parameter. Each element in the data set is assigned to a cluster center with the smallest distance to it. This method produces exactly *k* different clusters of greatest possible distinction. K-means cluster analysis is presented in Table 4.

Table 4: K-Means cluster analysis report

Variables	Between			Within		Prob
	DF1	DF2	Mean square	Mean square	F-ratio	Level
DFF	6	198	265.057	6.849812	38.70	0.000000
DM	6	198	284.5414	7.326143	38.84	0.000000
PH	6	198	1565.734	39.05067	40.09	0.000000
CPP	6	198	81.17728	1.236915	65.63	0.000000
PPC	6	198	10.87785	0.1932944	56.28	0.000000
PPP	6	198	1685.604	23.6639	71.23	0.000000
PL	6	198	15.58817	0.2645869	58.92	0.000000
SPP	6	198	61.50498	0.8602531	71.50	0.000000
TW	6	198	3.854053	0.2089823	18.44	0.000000
TP	6	198	135.2376	24.31667	5.56	0.000024
HI	6	198	861.5718	27.3555	31.50	0.000000
SCMR	6	198	1957.668	21.0882	92.83	0.000000
LWP	6	198	57.41039	0.6431783	89.26	0.000000
PC	6	198	47901.64	278.5432	171.97	0.000000
RWC	6	198	11220.73	100.2384	111.94	0.000000
SLA	6	198	79014.52	1934.228	40.85	0.000000
SYPP	6	198	168.7692	2.055387	82.11	0.000000

DFF: Days to 50% flowering; TP: Threshing %; DM : Days to maturity; HI : Harvest index (%); PH: Plant height (cm); SCMR : SPAD Chlorophyll meter reading; CPP : Cluster plant^{-1}; LWP : Leaf water potential (Mpa); PPC : Pods cluster^{-1}; PC : Proline content (µg g^{-1}); PPP: Pods plant^{-1}; RWC: Relative water content (%); PL : Pod length (cm); SLA : Specific leaf area; SPP : Seeds per pod; SYPP: Seed yield plant^{-1}; TW: test weight(g);

3.3. Clustering pattern and composition of group

Analysis revealed that a wide range of variability existed for all the traits studied indicating the presence of significant variation among the genotypes. Based on the K-means clustering analysis, all the 205 germplasm accessions including five check entries were grouped into seven different clusters as presented in the Table 5 and Figure 1. Cluster V was the largest with 38 genotypes followed by cluster I with 36, cluster III and VII with 28, cluster II with 28, cluster IV with 25 and cluster VI with 23 genotypes. The mode of distribution of genotypes from different geographical regions into various clusters was at random indicating that the genotypes originating from different agro-climatic / geographical regions grouped together into different clusters showing no parallelism between genetic diversity and geographical distribution. Our results are on par with findings of Raje and Rao et al. (2001), Venkateswarlu (2001), Dasgupta et al. (2005), Makeen et al. (2007), Tabasum et al. (2010), Divyaramakrishnan and

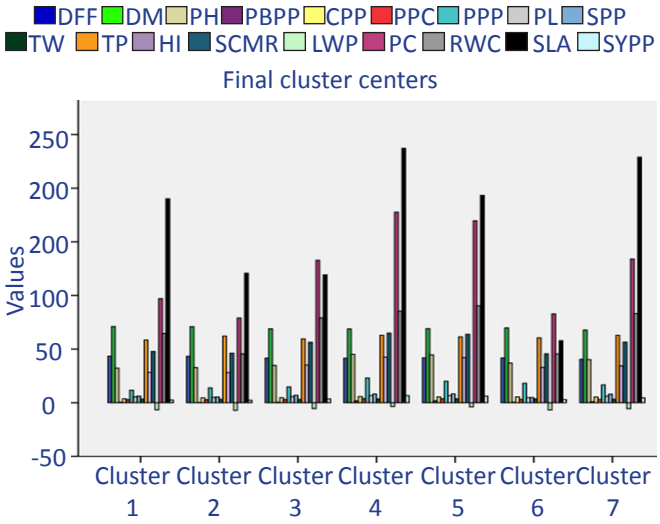

Figure 1: Bar diagram of distribution of characters into seven clusters

Table 5: Distribution of genotypes into 7 clusters as per K-Means Clustering

Clusters	No. of Genotypes	Genotypes
I	36	GG13-2, KM13-26, KM13-41, KM13-11, GG13-11, GG13-8, GG13-12, KM13-48, IPM2-3, IPM2-14, PDM-139, LGG-580, PM-112, LGG-578, LGG-563, LGG-594, TM-96-2, IC-546491, VC-6368, Harsha, LGG-460, KGS-5, Barimung-4, AKL-189, AKL-168, AKL-218, AKL-179, AKL-185, AKL-163, COGG-912, IC-73451, IC-105690, IC-73534, IC-73412, IC-39605, KKM-3
II	27	GG13-10, SML-668, LGG-596, LGG-572, LGG-583, AKL-106, AKL-225, AKL-95, AKL-194, AKL-195, LGG-593, PM-115, LGG-587, PM-113, IC-436775, IC-398984, IC-436767, IC-546481, VGG10-005, IC-39605, IC-73472, IC-92053, IC-73779, IC-73462, IC-53783, IC-73458, AKL-213
III	28	KM13-16,KM13-19,KM13-39, GG13-7, GG13-6, KM13-9, VGG07-003,AKL-39, KM13-55, GG13-4, LGG-595, AKL-182, IC-436646, IC-343964, IC-436528, IC-436723, VBN-1,KPS-1 , AKL-226, AKL-24, AKL-216, AKL-176, AKL-191,AKL-192, BGS-9,DGG-1,PDM 84-178, PS-16
IV	25	KM13-44, IPM99-125, AKL-103, AKL-211, KM13-11, KM13-30, KM13-45, KM13-18, KM13-5, KM13-02, KM13-12, LGG-586, IC-413311, IC-436767, VBN-223, COGG-912, VBN(G9)-3, ML-1165, VGG04-025, VGG04-004, VGG04-149, VGG-119, IC-118992, IC-73456, AKL-105
V	38	LGG-450, LGG-579, LGG-562, LGG-585, PLM-110, LGG-577, IC-436624, IC-413316, IC-436746, VGG10-010, VGG04-011, VGG04-007, COGG-93, VBNGG-2, TARM-2013, VGG04-005, COGG-920, VGG10-002, LGG-592, LGG-460, LGG-566, IC-436630, IC-436668, IC-436555, IC-413314, AKL-89, AKL-228, AKL-184, AKL-230, AKL-86, AKL-220, AKL-97, AKL-29, AKL-90, AKL-227, AKL-92, AKL-183, AKL-165
VI	23	LGG-590, AKL-212, KM13-37, KM13-23, GG13-9, KM13-49, GG13-54, KM13-20, GG13-5, Chinamung, KM13-47, KM13-11, IC-398988, COGG-954, VGG08-002, VC3890-A, DGGV-4, CGG-973, CN9-5, KPS-2, VC-6173, Harsha, MH-709
VII	28	LGG-588, LGG-589, LGG-582, AKL-170, IC-436723, VGG-112, IC-92048, LGG-591, LGG-584, LGG-555, LGG-564, IC-553514, IC-413319, IC-436542, IC-546493, IC-436594, AKL-20, AKL-229, AKL-169, AKL-84, AKL-82, AKL-174, AKL-161, AKL-180, AKL-222, AKL-187, AKL-200, AKL-164

Savithramma (2014), Suhel et al. (2015), John et al. (2015), Gunjeet et al. (2015), Muhammad et al. (2016) Wanga et al. (2017), Kaur et al. (2018), Sharma et al. (2018) and Mohan et al. (2019). Sanhita et al. (2019) reported formation of 4 clusters of mungbean genotypes for bruchid resistance when the data was subject to multivariate analysis.

3.4. Intra and inter cluster distances between clusters

The intra and inter cluster distances are presented in Table 6. The range of inter cluster distance was 45.01 to 208.17. The maximum inter cluster distance was recorded between the clusters I and VI (208.17) followed by cluster V and VI (168.52). The minimum inter cluster distance was recorded between the clusters IV and V (45.01) followed by cluster IV and VII (46.97). The range of intra cluster distance was 74.41 to 134.82. The maximum intra cluster distance was recorded for the cluster VI (208.17) followed by cluster IV (160.40). The minimum intra cluster distance was recorded for the cluster IV (45.01) followed by cluster V (52.55).

These results suggest that the genotypes grouped in different clusters may be used as potential parental lines for hybridization programmes to develop desirable genotypes as genetic diversity can be best exploited and chances of getting best transgressive segregants are more. The cluster means of 17 characters are presented in Table 6. From the data we can conclude that considerable variation exists for all the traits studied. Results showed that genotypes in Cluster V were early flowering (38.92 days) whereas genotypes in cluster VI were late flowering (47.00 days). The genotypes in cluster V were early maturing (66.24 days) whereas genotypes in cluster VI were late maturing (74.96 days). Cluster IV exhibited highest mean for plant height (45.70 cm) whereas the cluster IV showed lowest (25.62). Cluster per plant was highest in cluster III (6.95) and was lowest in cluster VI (1.01). Pods per cluster was highest in V (5.90) and lowest in cluster IV (2.37). Pods per plant was highest in cluster I (25.22) and was lowest in cluster IV (5.88). Pod length was highest in cluster V (6.74) and lowest in cluster VI (4.99). Seeds per pod was highest in cluster VII and was lowest in cluster IV (4.85). Test weight was highest in cluster IV (3.76) and lowest in cluster II (2.84). Threshing percentage was highest in cluster VI (63.82) and

Table 6: Intra and Inter cluster distances between cluster centers

Cluster	1	2	3	4	5	6	7
1		74.41	81.79	100.09	81.67	134.82	58.90
2	74.41		64.83	160.40	127.29	63.44	128.07
3	81.79	64.83		127.60	84.87	86.91	110.11
4	100.09	160.40	127.60		45.01	208.17	46.97
5	81.67	127.29	84.87	45.01		168.52	52.55
6	134.82	63.44	86.91	208.17	168.52		182.99
7	58.90	128.07	110.11	46.97	52.55	182.99	

lowest in cluster IV (57.90). Harvest index was highest in cluster V (41.93) and lowest in cluster VI (28.72).

The range of intra cluster distance was 74.41 to 134.82. The maximum intra cluster distance was recorded for the cluster VI (208.17) followed by cluster IV (160.40). The minimum intra cluster distance was recorded for the cluster IV (45.01) followed by cluster V (52.55). These results suggest that the genotypes grouped in different clusters may be used as potential parental lines for hybridization programmes to develop desirable genotypes as genetic diversity can be best exploited and chances of getting best transgressive segregants are more.

The cluster means of 17 characters are presented in Table 7. From the data we can conclude that considerable variation exists for all the traits studied. Results showed that genotypes in Cluster V were early flowering (38.92 days) whereas genotypes in cluster VI were late flowering (47.00 days). The genotypes in cluster V were early maturing (66.24 days)

Table 7: Cluster means and standard deviation values of traits in different clusters as per K-means Clustering

Sl. No.	Traits	Cluster-I		Cluster-II		Cluster-III		Cluster-IV		Cluster-V		Cluster-VI		Cluster-VII	
		$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	Σ	$\bar{X}$	σ	$\bar{X}$	σ
1.	DFF	40.67	2.86	43.56	2.64	39.11	2.13	42.32	3.20	38.92	2.02	47.00	2.34	45.21	3.05
2.	DM	68.64	2.88	70.96	2.75	66.46	2.40	70.08	3.28	66.24	2.15	74.96	2.36	72.61	3.08
3.	PH	41.89	4.45	30.95	7.19	37.39	7.14	25.62	5.24	45.70	6.01	33.59	8.92	43.89	4.61
4.	CPP	6.95	0.73	3.04	1.01	5.87	1.15	2.52	0.36	6.15	1.18	4.08	1.17	4.94	1.71
5.	PPC	3.62	0.44	2.79	0.49	2.91	0.42	2.37	0.40	3.90	0.41	2.77	0.56	3.86	0.37
6.	PPP	25.22	4.31	8.56	3.64	17.02	4.05	5.88	0.91	23.99	5.47	11.59	5.27	19.20	7.48
7.	PL	5.06	0.50	5.85	0.45	5.77	0.46	5.14	0.62	6.74	0.50	4.99	0.55	6.51	0.53
8.	SPP	5.15	1.17	6.74	0.92	6.87	1.23	4.85	0.69	8.23	0.73	5.68	0.95	8.43	0.57
9.	TW	3.47	0.41	2.84	0.53	3.37	0.53	3.76	0.59	3.74	0.35	2.89	0.49	3.54	0.30
10.	TP	61.73	3.82	60.99	5.53	58.06	7.16	57.90	3.46	62.77	5.14	63.82	2.54	61.58	5.15
11.	HI	31.35	5.13	31.36	5.83	34.40	4.78	30.63	6.53	41.93	4.88	28.72	5.61	41.35	3.84
12.	SCMR	47.15	3.90	54.02	5.99	53.98	4.76	46.00	4.72	65.06	4.17	44.82	3.73	62.19	4.75
13.	LWP	-6.61	0.58	-5.91	0.78	-5.75	0.78	-7.04	0.54	-3.88	0.95	-7.16	0.61	-3.86	1.13
14.	PC	85.19	15.24	122.34	23.74	124.16	20.71	85.06	12.02	173.03	15.97	77.24	10.08	168.03	14.46
15.	RWC	45.12	6.95	77.20	13.04	73.61	13.99	54.59	14.28	89.23	3.77	44.26	6.38	87.58	8.89
16.	SLA	90.93	49.90	167.00	52.05	150.27	45.87	114.72	51.28	210.61	36.52	109.30	44.53	219.37	21.51
17.	SYPP	4.51	1.36	1.54	0.48	3.89	1.20	1.05	0.20	7.38	1.82	1.89	0.98	5.76	2.38

$\bar{X}$ = Mean; σ =Std. deviation

whereas genotypes in cluster VI were late maturing (74.96 days). Cluster IV exhibited highest mean for plant height (45.70 cm) whereas the cluster IV showed lowest (25.62). Cluster plant^{-1} was highest in cluster III (6.95) and was lowest in cluster VI (1.01). Pods cluster^{-1} was highest in V (5.90) and lowest in cluster IV (2.37). Pods plant^{-1} was highest in cluster I (25.22) and was lowest in cluster IV (5.88). Pod length was highest in cluster V (6.74) and lowest in cluster VI (4.99). Seeds

pod^{-1} was highest in cluster VII and was lowest in cluster IV (4.85). Test weight was highest in cluster IV (3.76) and lowest in cluster II (2.84). Threshing percentage was highest in cluster VI (63.82) and lowest in cluster IV (57.90). Harvest index was highest in cluster V (41.93) and lowest in cluster VI (28.72).

Spad chlorophyll meter readings were highest in cluster V (65.06) and lowest in cluster VI (44.82). Leaf water potential was highest in cluster VII (-3.86) and lowest in cluster VI (-7.16). Proline content was highest in cluster III (124.16) and lowest in cluster VI (77.24). Relative water content was highest in cluster V (89.23) and lowest in cluster VI (44.26). Specific leaf area was highest in cluster VII (219.37) and lowest in cluster I (90.93). Seed yield plant^{-1} was highest in cluster V (7.38) and lowest in cluster IV (1.05). Three clusters namely; cluster IV, cluster V and cluster VI had maximum representation in terms of having either highest or lowest cluster means for the traits thus forming diverse group of genotypes. Umashankar and Sarkar (2018) have reported similar findings in green gram for the traits plant height, days to maturity, number of pods per plant, protein content and seed yield. Similar findings are also reported by Divyaramakrishnan and Savithramma (2014)

4. Conclusion

A wide range of variability was existed for all the traits studied indicating the presence of significant variation among the genotypes. Based on the K-means clustering analysis, all the 205 germplasm accessions including five check entries were grouped into seven different clusters The mode of distribution of genotypes from different geographical regions into various clusters was at random indicating that the genotypes originating from different agro-climatic / geographical regions grouped together into different clusters showing no parallelism between genetic diversity and geographical distribution.

5. Acknowledgement

Kanavi, M. S. P., thanks Director of Research, University of Agricultural Sciences, Bangalore for giving financial assistance to carry out the research work.

6. References

Azam, M., Hossain, M., Alam, M., Rahman, K., Hossain, M., 2018. Genetic variability, heritability and correlation path analysis in mungbean (*Vigna radiata* (L.)wilczek). Bangladesh Journal of Agricultural Research 43(3), 407−416.

Bangar, P., Chaudhury, A., Tiwari, B., Kumar, S., Kumari, R., Kangila, Bhat, K.V., 2019. Morphophysiological and biochemical response of mung bean [*Vigna radiata* (L.) Wilczek] varieties at different developmental stages under drought stress. Turkish Journal of Biology 43, 58−69.

Dasgupta, T., Mukherjee, K., Roychoudhury, B., Nath, D., 2005. Genetic diversity of horse gram germplasm. Legume Research 28(3), 166−171.

Divyaramakrishnan, C.K., D.L. Savithramma.,2014. Tailoring genetic diversity of mungbean [*Vigna radiata* (L). Wilczek] germplasm through principal component and cluster analysis for yield and yield related traits. International Journal of Agronomy and Agricultural Research (IJAAR) 5(2), 94−102.

Federer, W.T., 1956. *Augmented (or hoonuiaku) designs.* The Hawaiian Planters' Record. LV (2), 191-208.

Forgy, E.W., 1965. Cluster analysis of multivariate data efficiency vs. interpretability of classifications. Biometrics 21, 768−769.

Ghosh, S., Roy, A., Kundagrami, S., 2019. Diversity analysis of mungbean [*Vigna radiata* (L.) Wilczek] genotypes for bruchid resistance. Indian Journal of Agricultural Research 53(3), 309−314

John Kingsly, N.B., Packiaraj, D., Pandiyan, M., 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.

Kaur, G., Joshi, A., Jain, D., Choudhary, R., Vyas, D., 2015. Diversity analysis of green gram (*Vigna radiata* (L.) Wilczek) through morphological and molecular markers. Turkish Journal of Agriculture and Forestry 40, 229−240.

Kaur, G., Joshi, A., Jain, D., 2018. Marker assisted evaluation of genetic diversity in mung bean (*Vigna radiata* (L.) Wilcezk) genotypes. Brazilian Archives of Biology and Technology 61, e18160613.

Keatinge, J.D.H., Easdown, W.J., Yang, R.Y., Chadha, M.L., Shanmugasundaram, S., 2011. Overcoming chronic malnutrition in a future warming world: the key importance of mung bean and vegetable soybean. Euphytica 180, 129−141.

Kumar, V.U., Sarkar, K.K., Satish, G., Ramchander, 2017. Tailoring the genetic diversity of mung bean through cluster analysis for yield attributing traits to obtain efficient hybrids. Bulletin of Environment, Pharmacology and Life Sciences 6(3), 365−370.

Levene, H., 1960. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press, Palo Alto, 278−292.

Macqueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. Proceedings of 5th Berkeley Symposium. Mathematical statistics and probability. University of California Press 1, 281−297.

Makeen, K., Abrahim, G., Jan, A., Singh, A.K., 2007. Genetic variability and correlation studies on yield and its components in mung bean (*Vigna radiata* (L.) Wilezek). Journal of Agronomy 6, 216−218.

Mohan, S., Sheeba, A., Kalaimagal, T., 2019. Genetic Diversity and Association Studies in Greengram [Vigna Radiata (L.) Wilczek]. Legume Research, DOI: 10.18805/LR-4176, 1-6.

Muhammad, A., Muhammad, A.M., Qamar, U.Z., Muhammad, A.A., 2016. Uncovering the of mung bean (*Vigna radiata* L. Wilczek) genotypes under saline conditions using k-mean cluster analysis. Journal of Agriculture and Basic Sciences 1(1), 37−44.

Murthy, B.R., Arunachalam, V., 1966. The nature of genetic divergence in relation to breeding system in some crop plants. Indian Journal of Genetics and Plant Breeding 26, 188−198.

Panigrahi, K.K., Baisakh, B., 2014. Genetic diversity assessment for yield contributing characters of green gram [*Vigna radiata* (L.) Wilczek] cultivars from Odisha. Environment and Ecology 32(1A), 294−297.

Raje, R.S., Rao, S.K., 2001. Genetic diversity in a germplasm collection of mung bean (*Vigna radiata* (L.) Wilczek). Indian Journal of Genetics and Plant Breeding 61, 50−52.

Reddy, A.A., 2009. Pulses production technology: status and way forward. Economic and Political Weekly 44(52), 3−80.

Sharma, S.R., Singh, D., Kumar, P., Khedar, O.P., Varshnay, N., 2018. Assessment of genetic diversity in mungbean [*Vigna radiata* (L) wilczek] genotypes. International Journal of Genetics 10(7), 471−474.

Shusen, W., Alex, G., Michael, W.M., 2019. Scalable Kernel K-Means Clustering with Nystrom Approximation: Relative-Error Bounds. Journal of Machine Learning Research 20, 1−49.

Singh, R., Singh, M.K., Singh, A.K., Singh, C., 2018. Pulses production in India: issues and elucidations. Pharma. Innovation 7(1), 10−13.

Suhel, M., Singh, I.P., Bohra, A., Singh, C.M., 2015. Multivariate analysis in green gram [*Vigna radiata* (L.) Wilczek]. Legume Research 38(6), 758−762.

Tabasum, A., Saleem, M., Aziz, I., 2010. Genetic variability, trait association and path analysis of yield and yield components in mung bean (*Vigna radiata* (L.) Wilczek). Pakistan Journal of Botany 42, 915−3924.

Venkateswarlu, O., 2001. Genetic variability in green gram (*Vigna radiata* (L.) Wilczek). Legume Research 24, 69−70.

Wanga, L., Baib, P., Yuanc, X., Chena, H., Wanga, S., Chenc, X., Chenga, X., 2017. Genetic diversity assessment of a set of introduced mung bean accessions (*Vigna radiate* L.) The Crop Journal 6(2), 207−213.