Original Research Article

Molecular characterization and diversity analysis of green gram [*Vigna radiate* (L.) Wilczek] by using SSR markers

Abstract

Molecular characterization was done for greengram genotypes using SSR markers. The 14 green gram genotypes are clustered using ward method and squared Euclidean distance using nine SSR markers.As per dissimilarity coefficient, genotypes are grouped into three clusters namely Cluster I and Cluster II and Cluster III, Cluster I consists of 6 genotypes divides into two sub Clusters IA consist of IPM2-3,IPM-99-125,PUSA-0672,DGGV-2 and sub cluster IB consists of MH-2-5,HUM-16Cluster II consists of 4 genotypes divided into three sub clusters IIA consists of MH-3-18,PUSA VSHAL and sub cluster IIB consists of one genotype namely PANT MUNG-5, sub cluster IIC has PUSA-9531.Cluster III consist of 4 genotypes divided into two sub clusters IIIA consists of SML-668,SML-832 and sub cluster IIIB consist of IPMD-604-1-7,OBGG-58. Polymorphic information content values (PIC) ranged from 0.000 to 0.704 per locus with an average 0.519.Among 9 markers SSR VR91 generated highest percent of PIC value. Based on the polymorphic band results PCA was performed for 9 SSR markers for 14 greengram genotypes among all the factors PC_1 TO PC_9, the PC_1 (25.872) accounted for maximum proportion of variability in the set of all variables and PC_9(0.372) accounted for progressively lesser amount of variability.It can be concluded from the above results that the highest Eigen value for PC_1(2.328) and the least is PC_9 with (0.033).

Keywords: Molecular characterization, greengram, PIC, SSR markers

Introduction

Greengram [*Vigna radiata* (L.) Wilczek], also known as mungbean, is an ancient pulse crop widely cultivated in India. It is a diploid species with the chromosomal number 2n=2x=22, a member of the subfamily Papilionaceae of the Leguminosae family, and its botanical name is *Vigna radiata* (L.) Wilczek. Mungbean is a native of South Asia (India), and its likely ancestor is *Vigna radiatavar*. *sublobata*.Genetic diversity assessment in available cultivars has important implications in understanding the progress made in any breeding programme. Morphological markers are routinely used for estimating the genetic diversity, but recently molecular marker techniques have become powerful tools to analyze genetic relationships (Yoon *et al*., 2000 and Nath *et al*., 2017).

Molecular characterization is also useful in understanding the phylogenetic relationship between plant species to reveal the genetic diversity within a given taxonomic

group. Evaluation of genetic diversity would promote the efficient use of genetic variations in conservation (Paterson *et al*., 1991 and Das *et al*., 2014). The determination of purity of genotypes and utilization of germplasm in crop improvement are also possible with molecular markers (Samarajeewa*et al*., 2002 and Muthamilarasan and Prasad, 2015). DNA Fingerprinting is one of the important applications of molecular markers whichallows identifying an individual from others by using variations at DNA level.

Microsatellites or simple sequence repeats (SSRs) are clusters of short tandem repeated nucleotide bases distributed throughout the genome and are valuable for genetic mapping, genotyping and marker-assisted selection in breeding due to their co-dominant nature, higher allelic variation and even distribution in the genome (Gupta and Varshney, 2000 and Hernandez *et al*., 2002). Major features that made SSRs very popular are their abundant distribution in the genomes examined to date and their hyper variable nature (Toth *et al.,* 2000). In greengram, the variation observed for these markers is very less indicating the necessity to use more and more number of markers to get the clear picture of diversity and use it for fingerprinting of the genotypes.

Materials and Methods

Molecular Diversity in Greengram genotypes using SSR markerswas conducted during *Rabi* 2021-2022 at the field experimentation center and Seed Testing Laboratory of the Department of Genetics and Plant Breeding, Sam Higginbottom Institute of Agriculture, Technology and Sciences Deemed to be University, Allahabad followed by molecular work at Division of Plant biotechnology, Indian Institute Of Pulses Research (IIPR), Kanpur.The experimental material consisted of 14 diverse lines of greengram.

Genomic DNA extraction and quantification

DNA was extracted from leaf tissue collected at 21 days after sowing from all the entries using the method described by (Zheng *et al.,* 1995).The extraction buffer was prepared by using the chemicals (Table 1) and stock solutions as per the procedure.

Table 1. Chemicals required for preparation of extraction buffer

S. No.	components	stock concentration	working concentration	volume taken for 500 ml
1	Tris HCL	1M	200 mm	100 ml
2	EDTA	0.5M	20 mm	20 ml

3	NaCl	5M	1.4 mm	140 ml
4	CTAB	2%	2 g	10 g

SSR-PCR amplification

The polymerase chain reation (PCR) was carried out using a programmable thermocycler. 96 well PCR plates was taken and 2.5 µl of 20 ng template DNA was pipette into each of the PCR well after proper labeling and kept the PCR plate at 4^0C.

Chart 1.**Specification for polymerase chain reation**

Cycle	Denaturation		Annealing		Polymerization	
	Temp.	Time	Temp.	Time	Temp.	Time
First cycle	94˚C	5 min	-	-	-	-
39 cycles	94˚C	1 min	46-51˚C	2 min	72˚C	2 min
Last cycle	-	-	-	-	72˚C	10 min

Annealing temperature for each marker was standardized using gradient

The PCR products were analyzed by electrophoresis using a 3 % high resolution agarose gel using a gel Electrophoresis Unit.The DNA fragments were then visualized under UV-transilluminator and documented using gel documentation system.

Total of nine markers (Table 2) were used for characterization of 14 green gram genotypes. The amplification products were viewed under UV light and photographs were saved for the experimental evaluation. The amplification products were scored separately for each primer. The bands were scored for the presence or absence by binary coding i.e., assigning a value of 1 for presence and 0 for absence in a lane.

Table 2. Nine SSR Markers with sequence and Melting temperature (Tm) and Annealing temperature(AT)

No	Primers	Sequences	Tm	Length
1	VR 80	R=51AATGGTCCCTTTACCCCTTTT	50.5℃	21mer
		F=5'TGTGAGAGTGGAAGAGCAACTT	50.5℃	22mer
2	VR 48	R=5'AATAGGGCCCATAACATGTCC	52.4℃	21mer
		F=5'AGGTGAGTGAAAATTGGAATAGG	51.7℃	23mer
3	VR 91	R=5'TGGAGATGCAGGACTAAGAAGAG	55.3℃	23mer
		F=5'ACATATGTATCTGTCTGTGTGCCTA	54.4℃	25mer

4	VR 86	R=5'ATCGGTATATGTTGCCAATCAG	51.1℃	22mer
		F=5'CTATACTGCAATGAAGTGGATCTC	54.0℃	24mer
5	CEDG168	R=5'CATTACATTCCAGACCTGC	58.01℃	22mer
		F=5'CTGCTTGGTGTTGAGCTTC	57.8℃	25mer
6	CEDG030	R=5'TCGGCAGATAGAGGCTCACG	58.8℃	23mer
		F=5'TGAGGGAATGGGAGAGAGGC	58.6℃	22mer
7	CEDG050	R=5'GAGATTATCTTCTGGGCAGCAAGG	61.97℃	22mer
		F=5'TCCCACTTGTCCATTACCTCCAC	61.95℃	25mer
8	CEDG097	R=5'TGCCAAAGAGCCGTTAGTAGAA	58.21℃	23mer
		F=5'GTAAGCCGCATCCATAATTCCA	58.21℃	21mer
9	CEDG156	R=5'CTTAGTGTTGGGTTGGTCGTAAGG	61.97℃	21mer
		F=5'CGCGTATTGGTGACTAGGTATG	60.07℃	25mer

Data analysis

The amplification data obtained from SSR primers were used for a combined similarity matrix and to generate a combined dendrogram. Dendrogram of all test genotypes of greengram were constructed on the basis of presence or absence of SSR bands using the UPGMA (unweighted pair-group method with arithmetic mean) cluster analysis.Principal component analysis (PCA) was also usedfor nonhierarchical relationships among the genotypes.Eigenvalues and eigenvectors were calculated by the Eigenprogram using a correlation matrix as input (calculatedusing standardized molecular data), and 2D and 3Dplots were used to generate the two-dimensional PCA plotfrom XLstat version 2021.

Polymorphic information content

Polymorphic information content (PIC) that provides an estimate of the discriminatory power of a locus or loci, by taking into account, was calculated using the formula given by **Smith et al., 1997.**

$$PIC = 1 - \sum_{f=1}^{n} Pij^2$$

Where P*ij*is the frequency of the jth allele for the ith marker, and is summed over n alleles. The calculation was based on the number of alleles per locus.

Results and discussion

SSR analysis

The present study evaluated the genetic diversity of 14 greengram genotypes.The studyof genetic diversity in any breeding population is essentialas it constitutes the backbone of any breeding and improvementprogram. It helps in the development of crop that issuitable and adaptable to rapid climate change through theintroduction of foreign genes.Thus, genetic diversityis needed for developing ideal and desired crop varieties forpresent and future needs. In this study, 30 SSR markers were screened, out of which 9 SSR markers were found to be polymorphic and suitablefor diversity analysis. The use of SSR for greengram diversity studyis very crucial as it provides accurate and unbiased assessment and reveals in depth information on the genetic divergence of the material. SSR marker has been widely recognizedfor its codominant inheritance pattern, high informativepower, and transferability among the species, hence,its superiority as a marker of choice for plant improvement program.

From 9 SSR markers screened, 8 markers displayed clear and repeatablepolymorphic bands, were selected for analysis as shown infigures 1 and 2. A total of 20 alleles wererecorded, andthenumber of alleles per locus rangedfrom2inVr80, Vr86, CEDG 168, CEDG 030, CEDG 50 and CEDG 156 to 5 in Vr91 with an average of 2.22. The expectedheterozygosity differed among the markers and it rangedfrom 0.459 (Vr80) to 0.745 (Vr91) with an average of 0.60.

The polymorphic information content (PIC) was used as a measure to know the polymorphism of a marker locus and genetic diversity evolution. The PIC values were varied and are considered as informative when the values are more than 0.50 and PIC values of less than 0.25 are considered as low level of diversity. PIC values will be moderately informative if values are 0.25 to 0.50. PIC was observed to differs significantly, from 0.354 (Vr80) to 0.704 (Vr91). Similar results were also reported by Kushwaha *et al*. (2013), Andeden*et al*. (2013), Karthikeyan *et al*. (2012) and Suvan *et al*. (2020) for PIC values in the range of medium to high, while Gopal Krishna *et al*. (2010) and Kumar *et al*. (2017) reported low PIC values. This high PIC marker can be used for diversity studies, gene mapping and to know germplasm evolution.

TABLE 3: Information generated by using8 SSR Markers in Greengramgenotypes

Markers	Alleles count	No. of effective alleles	Observed heterozygosity	Expected heterozygosity	PIC	I
Vr80	2.00	1.85	0.00	0.459	0.354	0.652
Vr48	3.00	2.09	0.00	0.520	0 .464	0.892

Vr91	5.00	3.92	0.00	0.745	0 .704	1.470
Vr86	2.00	1.96	0.00	0.490	0.370	0.683
CEDG 168	2.00	2.00	1.00	0.500	0.375	0.693
CEDG 030	2.00	2.00	1.00	0.500	0.375	0.693
CEDG 50	2.00	2.00	1.00	0.500	0.375	0.693
CEDG 097	0.00	0.00	0.00	0.000	0 .000	0.000
CEDG 156	2.00	2.00	1.00	0.500	0.375	0.693
Total	20.00	15.81	4.00	4.214	3.392	5.776
Mean	2.22	1.98	0.44	0.468	0.519	0.641

PIC: Polymorphism information content **I :**Shannon's information diversity index

Shannon's information index(I) is used to randomly measure the diversity among the species Shannon (1948) in which the present study given the result high (I) diversity result 1.470 for VR91 and lowest diversity 0.000 for CEDG097 and the average Shannon's index is 0.641 for the 9 SSR polymorphic markers as shown in table 3.

Results from the present investigation revealed remarkablyabundant genetic variation among the 14 greengramgenotypes.Thenumber of alleles ranged from 2 to 5.The number of alleles indicates the richness of thepopulation. Since SSR are short tandem repeats, generally allele numbers of 2 to 5 alleles per locus are considered goodas seen in this study. The PIC value ranged from 0.354 to 0.704 with an average of 0.519. The richness of information amarker can give, otherwise known as PIC reported in thisstudy, was very interesting.

High PIC of 0.354 to 0.704 as seen in this study revealedthat the markers have the required properties to be used indiversity study.Genetic diversity indices such as expectedheterozygosity as well as Shannon's index among the markers were very high, thus reflecting the heterozygous nature of the population.

The average number of alleles per locus (Na) varied from 2.00 to 5.00. The effective number of allelesper locus (Ne) was small compared to the number of allelesper locus and it ranged from 1.849 to 3.920 with the mean number of 1.979. Shannon's information index in the present studywas anotherindication of the presence of high genetic diversity in the greengram genotypes under consideration.

Vr 80 @ 51°C WITH G.G DNA ,LADDER=50BP(LADDER=GENEI)

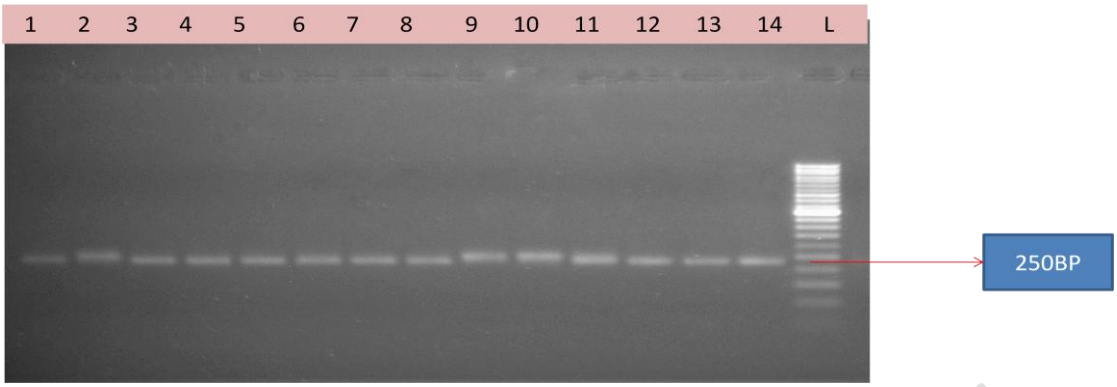

Figure 1: Image of DNA Banding pattern generated by Agarose gel electrophoresis of SSR VR 80 marker for 14 green gram genotypes.

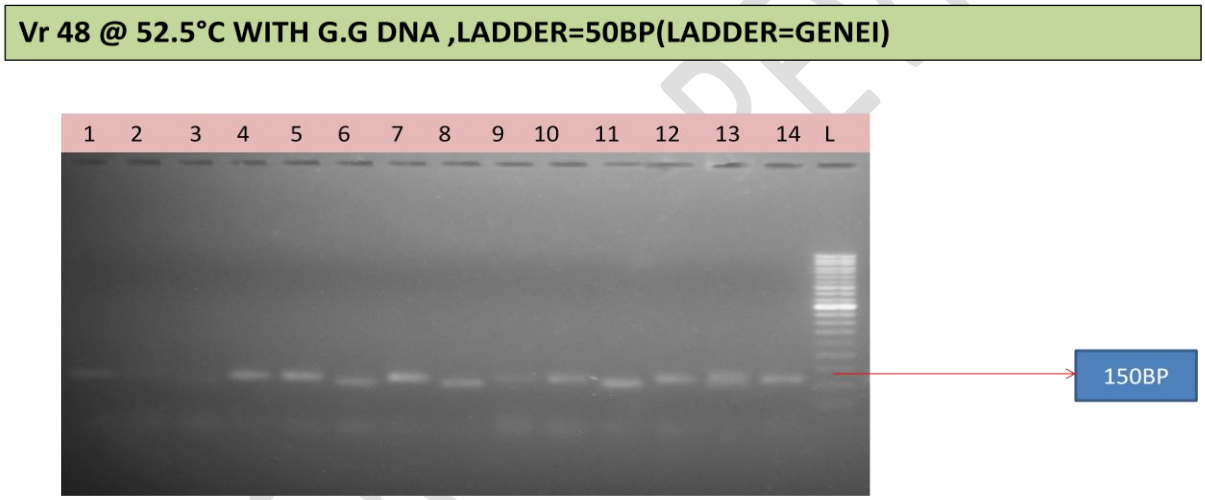

Figure 2: Image of DNA Banding pattern generated by Agarose gel electrophoresis of SSR VR 48 marker for 14 green gram genotypes

Principal component analysis

Principal coordinate analysis (PCoA/ PCA) for morphological (visuallyassessed and measurable characters) and molecular markers were carried out tocorroborate and validate the pattern of cluster analysis based on UPGMA method. PCAis utilized to drive a 2-dimensional and 3- dimensional scatter plot of individuals, suchthat geometrical distances among individuals in the plot reflect the genetic distancesamong them with minimal distortion.

In the present study, selected 14 genotypes werecharacterized according to 9groups by PCA analysis. Total variation was divided in to 9 principal components. The first four PCs, hadeigen values more than 1 and explained the 75.458%variation cumulatively were selected as these attributescontribute more to variation than remaining ones (Figure 3).

Eigenvalues, variability %, and cumulativecontribution of each component to variation alongwith the contribution of each character to respectivePCs are presented in Table 4. PC1 explained 25.872%of the total variability and PC2 with 19.644 % contribution was highest. Factor loading of principal components and factor scores principal components of 14 genotypes of greengram genotypes are presented in table 5 and 6 respectively.

Biplot analysis

To check the diversity among selected genotypesthey were plotted on biplot regarding the first twoPCs that had Eigenvalue greater andcontributing 45.51% variability. Genotypes that areclosely located on biplot, perceived as alike when ratedon given attributes. More the distance between thepoint of origin and genotype, more diverse will be thegenotypes from others. Regarding the first two PCs,genotypes were differentiated into four diverse groups (Figure 4). On biplot, three genotypes i.e. PUSA VISHAAL, IIPM-99-125 and IPM-2-3clogged far away from the originand were considered as diverse from the others (Figure 4). Remaining genotypes were clustered intothree groups. Genotypes i.e. DGGV-2, PUSA 0672, MH-2-5 and IPMD-604-1-7were clustered togetherand closer to origin as well hence, these genotypes areless diverse and have less breeding value. Genotypesi.e. SML-832, SML-668 and OBGG-58 formed the third group while genotypesPANT MUNGS-5, PUSA 9531, MH-3-18 and HUM-16 were differentiated from rest of thegenotypes and clustered in fourth group.

TABLE 4: Eigen values, %variance and cumulative eigen values of promising lines of greengram

Observations	Eigenvalue	Variability (%)	Cumulative %
P1	2.328	25.872	25.872
P2	1.768	19.644	45.516
P3	1.587	17.632	63.148
P4	1.108	12.310	75.458
P5	0.873	9.699	85.157

P6	0.716	7.954	93.111
P7	0.453	5.030	98.141
P8	0.134	1.487	99.628
P9	0.033	0.372	100.000

Table 5: Factor loading of principal components in Greengram

Observations	P1	P2	P3	P4	P5	P6	P7	P8	P9
Vr80	0.671	0.243	-0.199	0.095	0.294	-0.530	0.268	-0.033	0.052
Vr48	0.198	0.889	0.171	-0.193	-0.046	0.230	-0.192	0.018	0.104
Vr91	-0.346	-0.114	0.647	-0.382	-0.435	-0.209	0.229	-0.124	0.035
Vr86	0.594	-0.029	-0.267	-0.506	-0.012	0.433	0.360	0.013	-0.027
CEDG 168	0.690	-0.495	0.175	0.349	-0.011	0.251	-0.116	-0.218	0.042
CEDG 030	0.280	0.357	0.765	-0.121	0.421	-0.015	-0.064	-0.053	-0.097
CEDG 50	0.586	-0.570	0.490	-0.121	-0.045	-0.078	-0.100	0.241	0.036
CEDG 097	-0.644	-0.184	0.314	0.231	0.499	0.263	0.274	0.041	0.066
CEDG 156	0.231	0.416	0.271	0.674	-0.408	0.110	0.244	0.082	-0.034

Table 6: Factor scores principal components of 14 genotypes of greengram genotypes

Observations	P1	P2	P3	P4	P5	P6	P7	P8	P9
PUSA VISHAAL	3.163	0.748	-1.418	-0.872	-1.853	-0.857	-0.694	-0.107	-0.058
IIPM-99-125	2.296	0.815	-0.057	-0.04	1.013	0.329	0.174	0.312	0.085
SML-832	-2.101	0.861	-1.479	0.484	-0.591	0.156	-0.246	0.124	0.474
DGGV-2	0.032	-0.061	1.196	0.526	0.442	0.436	-1.126	0.14	-0.041
IPMD-604-1-7	-0.176	0.462	-1.443	0.247	0.103	1.875	0.204	-0.774	-0.041
PANT MUNGS-5	-0.033	-2.38	-0.938	0.154	-0.791	0.801	0.978	0.357	-0.234
SML-668	-2.592	1.565	-0.913	-0.178	-0.314	-0.33	0.006	0.645	-0.253
PUSA 9531	-0.963	-2.736	-0.387	-0.315	0.567	-0.523	-0.944	-0.317	-0.046
OBGG-58	-0.788	1.904	-0.066	0.25	1.066	-1.282	0.461	-0.552	-0.222

IPM-2-3	1.924	0.4	0.129	0.126	1.301	0.361	0.629	0.304	0.089
MH-3-18	0.269	-1.703	-0.269	1.206	0.314	-1.558	0.642	-0.038	0.17
PUSA 0672	0.064	0.004	1.24	0.66	0.352	0.463	-1.05	0.186	-0.079
HUM-16	-0.889	-0.313	1.395	-3.307	0.147	0.032	0.385	-0.059	0.124
MH-2-5	-0.206	0.434	3.01	1.058	-1.757	0.098	0.58	-0.222	0.03

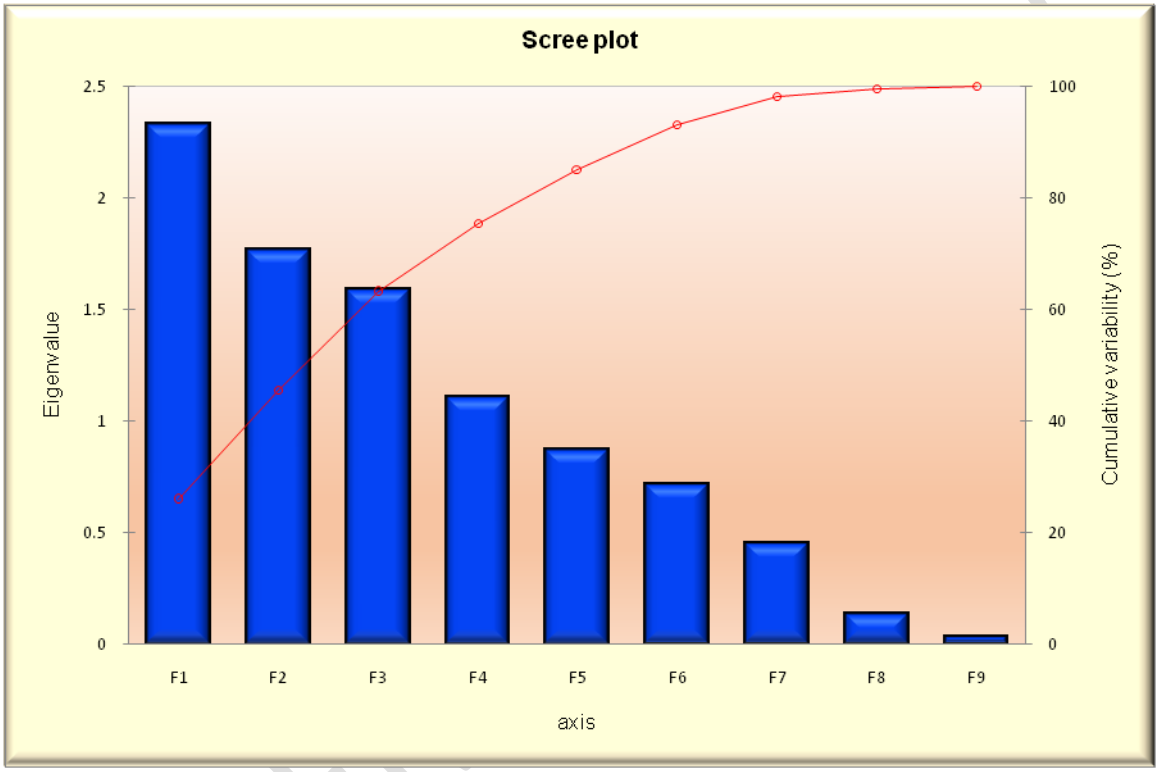

Fig. 3: Screen plot showing PCA factors eigen value and cumulative variability

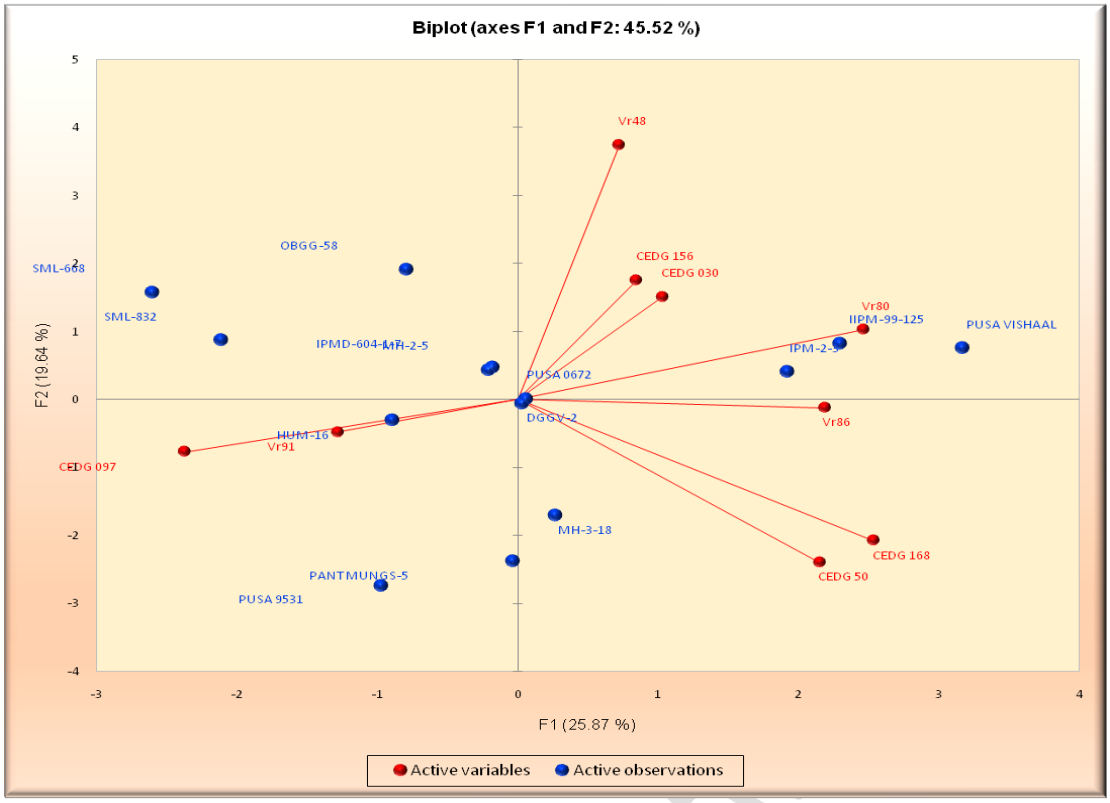

Fig. 4: Distribution of 20 genotypes on the bi-plot axes in PC 1 and PC2 of principal component analysis

Cluster Analysis

The 14 green gram genotypes are clustered using ward method and squared Euclidean distance using 9 SSR markers. Dendrogram showing genetic relationship among 14 greengram genotypes based on Euclidean's coefficient and UPGMA analysis using molecular characterization of the aforesaid loci and designated into three major clusters (Figure 5). As per dissimilarity coefficient, genotypes are grouped into three clusters namely Cluster I and Cluster II and Cluster III.

Cluster I consists of 6 genotypes divides into two sub Clusters IA consist of IPM2-3,IPM-99-125,PUSA-0672,DGGV-2 and sub cluster IB consists of MH-2-5,HUM-16. Cluster II consists of 4 genotypes divided into three sub clusters IIA consists of MH-3-18,PUSA VSHAL and sub cluster IIB consists of one genotype namely PANT MUNG-5, sub cluster IIC has PUSA-9531.Cluster III consist of 4 genotypes divided into two sub clusters IIIA consists of SML-668,SML-832 and sub cluster IIIB consist of IPMD-604-1-7 and OBGG-58. These results were in agreement with the reports Sharma and Sirohi (2018), Kanavi *et al.* (2019) and Tabasum *et al.* (2020).

Conclusion

No variation was observed between the clusters generated through Euclideandissimilarity and PCA analysis for measurable characters. The major clusters generatedin both the cases were same.PCA gave a better defined structure than the UPGMA dendrogram since PCAused two and three dimensions against only one dimension in case of dendrogram. Thebetter presentation using PCA was observed for the difference in relatedness of somegenotypes.

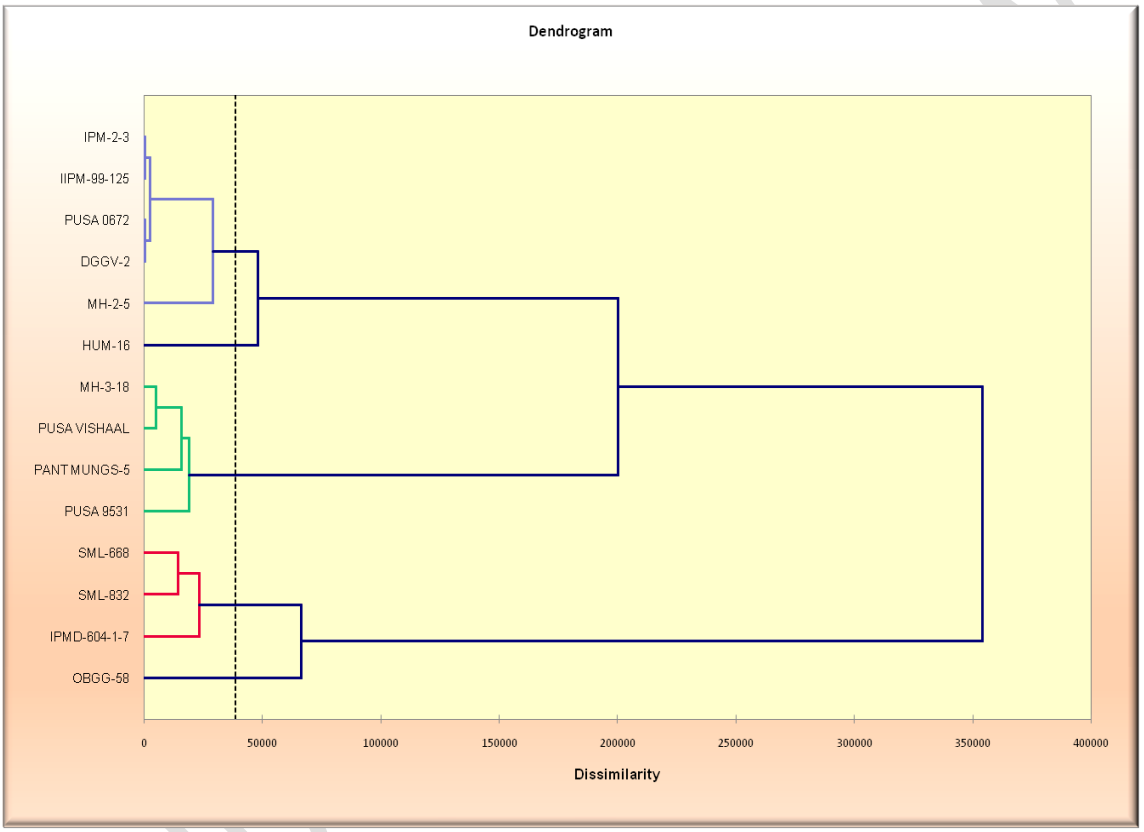

Fig. 5: Cluster analysis using ward method and squared Euclidean distance for 14 Greengram genotypes.

References

Das, S., Das, S.S and Ghosh, P. (2014). Assessment of molecular genetic diversity in some green gram cultivars as revealed by ISSR analysis. *Advances in Applied Science Research.* 5(2): 93-97.

Gupta, P and Varshney, R. (2000). The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. *Euphytica.*113(3): 163–185.

Hernandez, P., Lauri, D. A., Martin, A and Snape, J. W. (2002). Utility of barley and wheat simple sequence repeat (SSR) markers for genetic analysis of *Hordeum chilense*and titrodeum. *Theoretical and Applied Genetics.* 104: 735–739. 10.1007/s001220100674.

Kanavi, M.S.P., Rangaiah, S and Krishnaprasad, B.T. (2019). Molecular characterization and genetic diversity analysis through SSR markers in germplasm lines of green gram [*Vigna radiate* (L.) Wilczek]. *International Journal of Bio-resource and Stress Management*. 10(3): 298-305.

Muthamilarasan, M and Prasad, M. (2015). Advances in *Setaria*genomics for genetic improvement of cereals and bioenergy grasses. *Theoretical and Applied Genetics*. 128: 1–14.

Nath, A., Maloo, S.R., Bharti, B., Dubey, R.B and Verma, R. (2017). Combining ability analysis in mungbean [*Vigna radiata* (L.) Wilczek]. *Legume Research*. 41: 519-525.

Paterson, A.H., Damon, S., Hewitt, J.D., Zamir, S., Rabinowitch, H.D., Lincoin, S.E., Lander, S.E and Tanksley, S.D. (1991). Mendelian factors underlying quantitative traits in tomato: comparison across species, generations, and environments. *Genetics Society of America*. 127: 181-197.

Samarajeewa, P.K., Nanayakkara, H.L.V., Ekanayake, E.M.D.S.N and Sumanasinghe, V.A. (2002). RAPD analysis of genetic relationships of wild and cultivated *Vigna* species. *Annals of Sri Lanka Department of Agriculture.* 4: 215-226.

Sharma, S and Sirohi, A. (2018). Estimation of genetic diversity of mungbean (*Vigna radiata* L.) cultivar using SSR and ISSR marker analysis. *Journal of Pharmacognosy and Phytochemistry.* 1: 207-211.

Smith J.S.C., Chin E.C.L., Shu H., Smith O.S., Wall S.J., Senior M.L., Mitchell S.E., Kresovich, S and Zeigle J. (1997). An evaluation of the utilityof SSR loci as molecular markers in maize (*Zea mays* L.):comparisons with data from RFLPs and pedigree. Theor ApplGenet 95: 163–173.

Tabasum, A., Hameed, A and Asghar, M.J. (2020). Exploring the genetic divergence in mungbean (*Vigna radiata* L.) germplasm using multiple molecular marker systems. *Molecular Biotechnology*. 62: 547–556.

Toth, G., Gaspari, Z and Jurka, J. (2000). Microsatellites in different eukaryotic genomes survey and analysis. *Genome Research*. 10(**7**): 967-81. doi: 10.1101/gr.10.7.967.

Yoon, M.S., Doi, K., Kaga, A., Tomooka, N and Vanghan, D.A. (2000). Analysis of genetic diversity in the Vigna minima complex and related species in East Asia. Journal of Plant Research. 113: 375-386.

Zheng, K., Subudhi, P.K., Domingo, J., Magantay, G and Huang, N. 1995. Rapid DNA isolation for marker assisted selection in rice breeding. *Rice Genetics NewsLetter*. 12:255-258.