Received: 16 Jul 2022 | Revised: 21 Dec 2022 | Accepted: 10 Jan 2023

Electronic Journal of Plant Breeding

Research Article

Genetic diversity study through K-Means clustering in germplasm accessions of chickpea (*Cicer arietinum L.*)

V. Jayalakshmi*, A. Lokeshwar Reddy, S. Rama Devi and M. Mohammed Imran

Regional Agricultural Research Station, Acharya N.G. Ranga Agricultural University, Nandyal,
Andhra Pradesh-518502.
***E-Mail** : v.jayalakshmi@angrau.ac.in

Abstract

To identify potential genotypes for location specific breeding programmes of Andhra Pradesh, genetic diversity among 537 chickpea germplasm accessions was assessed for yield attributing traits during *rabi* 2020-21 at Regional Agricultural Research Station, Acharya N.G. Ranga Agricultural University, Nandyal, Andhra Pradesh. K-means clustering was utilised to ascertain genetic distances among 537 chickpea germplasm accessions which were grouped into 12 different clusters. Cluster I was the biggest and had the most genotypes (73), followed by clusters XII (66), VII (60), XI (58), X (49), V & VII (43), III (40), IV (36), II (27), IX (22) and VI (20). Clusters IV and IX, as well as clusters VI and VII, were the most divergent, with the maximum inter-cluster distances. The germplasm accessions *viz.,* NBeG-169, RVSSG-40, CP 117-5-23 and ICCV 07101 were identified with high *per se* performance from diverse clusters IV, IX, VI, and VII for inclusion as parents in a hybridization program to achieve desirable transgressive segregants.

Keywords: Chickpea, cluster, genetic diversity and K-means clustering

Chickpea is the only cultivated species in the genus *Cicer*, with 2n=2x=16 chromosomes and a small genome size of 738.09 Mbp (Varshney *et al.,* 2013). It is a member of the sub-family Papilionaceae of the family Leguminosae. The genus includes 43 species, nine of which are annual, 33 of which are perennial and one of which is unclassified. It is an important cool season grain legume with an indeterminate growth habit that is widely cultivated throughout the world's tropical and subtropical regions (Gaur *et al.,* 2012) in more than fifty countries and regions including Central-West Asia, Southern Europe, Ethiopia, North Africa, and Australia. According to Van der Maesen (1972), the species was originated in the Southern Caucasus and Northern Persia.

India contributes 70 % of total global chickpea production with 116.2 lakh tonnes grown on 112 lakh hectares with a productivity of 1036 kg/hectare (www.agricoop.nic.in). India is the world's largest producer of gram, followed by Turkey, Pakistan, Myanmar, and Ethiopia. Chickpea ranks first in total pulse production in India, followed by redgram. Andhra Pradesh is sixth important chickpea growing state in India and produces 5.66 lakh tonnes on an area of 4.65 lakh hectares with a productivity of 1218 kg/hectare. (Annual report, All India Co-ordinated Research Project on Chickpea, 2020-21). The typical climate in Andhra Pradesh with short and warm winters and intermittent dry spells during growing season of chickpea warrants the need for specialised breeding programmes. One of the critical criteria for selecting parents in the crop breeding is to isolate best genotypes from transgressive segregants. Cluster analysis would definitely help plant breeders in identifying genetically diverse parents who fall into different clusters. Clustering is a technique that groups together millions of data points to form a cluster. Cluster analysis or clustering is the process of grouping, categorizing or classifying a set of objects into many subsets called clusters in such a way that items within one subset are more "similar" to each other, while items within other subsets are "dissimilar." As a result, there must be a way to differentiate between "dissimilar" and "similar" items. K-means clustering is a fundamental clustering technique that is used to analyze data points. K-means is a most widely used algorithm for clustering with known sets of median points. Clustering can be used to discover meaningful groupings within a data set in an exploratory manner or it can serve as the starting point for more advanced analysis (Wanga *et al.,* 2017). K-means clustering is a machine learning algorithm that is hearsay because it is a powerful genetic diversity assessment technique that generates genetically diverse clusters / heterotic groups based on genetic distances between germplasm accessions. Once the heterotic groups are created, then it is simple to identify clusters that are genetically distant and the germplasm accessions that fall into these clusters are also genetically diverse. Plant breeders will find it easier to identify genetically diverse germplasm accessions, which will then serve as parental lines in a crossing program. Present research was carried out with a purpose to identify genetically diverse genotypes from 537 *desi* chickpea germplasm which can be later used as parental lines in plant breeding programmes to develop genotypes specifically suited to Andhra Pradesh.

The experiment was conducted at Regional Agricultural Research Station, Acharya N.G. Ranga Agricultural University, Nandyal, Andhra Pradesh with 537 chickpea germplasm accessions. The genotypes consisted of germplasm lines collected from ICRISAT, Patancheru, Indian Institute of Pulses Research, Kanpur and also lines obtained from State Agricultural Universities involved in chickpea breeding programmes. One hundred and eighty six advanced breeding lines developed at RARS, Nandyal, Andhra Pradesh were also included in the study. The experiment was conducted during *rabi* 2020-21 in Alpha Lattice with two replications. Each genotype was sown in single row in a plot of 4 m length at a spacing of 30 cm between rows and 10 cm between plants within row. For all characters except days to 50 percent flowering and days to maturity, observations were made on five randomly selected competitive plants from each germplasm accession. The mean of the respective characters was calculated by averaging the values of five competitive plants. Days to 50 per cent flowering, days to maturity, number of primary branches, plant height (cm), number of pods, 100-seed weight (g) and seed yield per plant (g) were recorded. The data was subjected to analysis of variance (ANOVA). The germplasm accessions were classified using the 'K-means clustering' model proposed by Macqueen (1967) and Forgy (1965). OPSTAT statistical software was used to perform K-means clustering analysis.

Analysis of variance revealed significant genotypic difference in 537 genotypes for all seven quantitative traits under study. Genetic diversity among the parents is a prerequisite for ensuring the chances of selecting improved segregants for various characters (Vishnu *et al.,* 2018). The success of hybridization in highly self-pollinated crops like chickpea mainly depends on the genetic diversity of the parents. K-means clustering was found to be potential to quantifying the degree of divergence in germplasm collections of crop plants. Clustering of 537 chickpea germplasm accessions using K-means clustering grouped them into 12 clusters (**Table 1**). Cluster 1 consisted of maximum number of genotypes (73), followed by cluster 12 (66), cluster 7 (60), cluster 11 (58), cluster 10 (49), clusters 5 and 8 (each with 43), cluster 3 (40), cluster 4 (36), cluster 2 (27), cluster 9 (22) and cluster 6 (20). The distribution of genotypes from different agro-climatic regions grouped together into different clusters, indicating that there was no parallelism between genetic diversity and geographical distribution. Similar observations were made by Jeethava *et al.* (2000), Raval and Dobariya (2004), Jayalakshmi and Ronald (2011), Parameshwarappa *et al.* (2011), Parashi *et al*. (2013), Jayalakshmi *et al.* (2014), Kuldeep *et al.* (2015) and Jayalakshmi *et al.* (2016).

The inter cluster distances were measured to identify genetically diverse clusters. It was ranged from 10.33 to 39.55 (**Table 2**). Minimum inter cluster distance was observed between clusters III and XI (10.33), followed by clusters XI and XII (10.41). The shortest inter cluster distance indicated that the genotypes contained in these clusters were less diverse. It indicated a close relationship and similarity among the genotypes for the majority of the characters. Maximum inter cluster distance was observed between clusters IV and IX (39.55), followed by clusters VI and VII (36.52) implying that the genotypes of these clusters were genetically diverse. Hybridization between the genotypes of the clusters IV and IX, VI and VII might result in high heterotic expression for yield components (Sreelakshmi *et al.,* 2010).

The value of clusters for selecting genotypes for various attributes was evident from cluster mean value. A perusal of the results on cluster means for yield attributing traits (**Table 3**) revealed considerable differences between the clusters for all characters under study. The genotypes of cluster VIII exhibited early flowering (40.16 days) and matured quickly (76.28 days), whereas cluster II genotypes exhibited late flowering (60.30 days) and matured slowly (95.26 days). Cluster I had the highest mean number of primary branches (4.23), while cluster III had the lowest (3.24). Cluster VI had the highest plant height (48.97 cm) and cluster III had the lowest (32.84 cm). Cluster IX had highest number of pods (46.18), while cluster IV had lowest (10.50).Cluster IX had the highest seed yield per plant (7.59 g), while cluster IV had the lowest (3.15 g). Cluster III had the highest 100 seed weight (33.59 g) and cluster IX had the lowest (19.09 g). The genotypes of

Table1. Distribution of promising genotypes in diverse clusters for various attributes as per K-Means clustering

Cluster number	Number of genotypes	Days to 50% flowering	Days to maturity	Number of primary branches	Plant height	Number of pods	100 seed weight	Seed yield per plant
I	73	ICCIL 04023 (36), BCP-60 (36), B-48 (37), BBG-3 (37)	NBeG 158 (71), BBG-3 (72), BDNG-804 (73),	NBeG 49 (7), ICCV 09106 (7), ICCV 09109 (7), ICCV 11111 (7)	NBeG 781 (45), BG-3024 (48.2), NBeG 798 (49.2), PG-0809 (50.5), ICCIL 85213 (53)	NBeG 699 (38), NBeG 158 (39.5)	ICCV 07107 (32.2), ICCV 09106 (32.3), ICCV 06107 (33.5), NBeG 47 (35.5), BG-3061 (35.8)	NBeG 690 (9.8), BCP-60 (9.8), NBeG 816 (9.9) BBG-3 (10.1), B-48 (10.2), BDNG-804 (10.7), NBeG 1004 (10.8), NBeG 798 (11), BG-256 (11), RVSSG-35 (12.9), ICCV 07107 (16.7)
II	27			CP-116-14-21-P3 (6)	CP-116-14-21-P1 (52.7), RKG-160 (53), CP-116-15-24 (53.3), CP-232-6-4 (53.5), CP-117-10-10 (55.3),			
III	40	NBeG 538 (40), NBeG 95 (44), NBeG 475 (44), ICCV 08106 (44), ICCV 11101 (44)	ICCV 08110 (82), ICCV 08101 (84), ICCV 08111 (84)	ICCV 07116 (6), ICCV 10111 (6.5), ICCV 05115 (7.5)			ICCV 08106 (38.5), NBeG 393 (39.5), NBeG 25 (39.8), NBeG 181 (43.6)	ICCV 08111 (10), ICCV 07108 (10.5), ICCV 07115 (14.1)
IV	36				CP-232-6-12 (50.9), IPC-08-11 (51.9), IPC-2010-134 (52.2), CP-116-14-23 (52.5), CP-117-5-1 (56.4)		BGD-1068 (29.7), BGM-572 (29.7), IPC-06-119 (29.7), CP-117-5-16 (29.7), IPC-06-64 (29.7)	
V	43	ICCV 12118 (35), ICCV 15116 (36), ICCV 13108(37), JG-226(37), RBC-37 (37)	RBC-37, RVSSG-66 (73), ICCV 1510175), ICCV 15112 (75)	ICCV 13640 (6), ICCV 15112 (6), ICCV 08102 6.5), ICCV 15104 (6.5)	ICC 9862 (53), ICCV 13608 (53.4), Radhey (57), ICCV 13640 (57.2), ICCV 13647 (60.7)		ICCV 12102 (35), ICCV 15118 (35.8), GJG-0920 (36.3), NBeG 786 (38.4), ICCV-03105 (43.4)	
VI	20			CP-116-15-4-P2 (6.5)	CP-116-15-4-P4 (52.7), CP-117-11-9 (53.7), ICC 10945 (54), CP-117-11-10 (54.1), CP-116-14-1 (58.1)	ICC 10939 (43.7), CP-211-2-16 (43.7), CP-116-14-17 (45.5), CP-117-5-23 (45.5), BDNG-2010-1 (63.5)		CP-117-5-23 (13.5)

Table 1. Continud..

Cluster number	Number of genotypes	Days to 50% flowering	Days to maturity	Number of primary branches	Plant height	Number of pods	100 seed weight	Seed yield per plant
VII	60	NBeG 829 (34), NBeG 637 (35), NBeG 625 (36), ICCV 10107 (36)	NBeG 157 (73), ICCV 15102 (74), ICCIL 04022 (75), NBeG 648 (76)	ICCV 10113 (6), ICCV 11102 (6.5), ICCV 11106 (8), ICCV 07104 (10)			ICCV 10107 (29.9), NBeG 584 (31), NBeG 637 (31), NBeG 555 (33.3), NBeG 625 (34.1),	ICCV 07101 (11.3)
VIII	43	NBeG 633 (34), NBeG 779 (34), NBeG 776 (35), NBeG 778 (36)	ICCV 11112 (70), NBeG 815 (70), NBeG 620 (72), NBeG 738 (72), NBeG 740 (72)	NBeG 280 (7), NBeG 652 (7)			NBeG 773 (37.1), ICCV 11117 (37.9), NBeG 631 (38.1), NBeG 815 (41.4), NBeG 155 (45.7)	
IX	22	AKG-9826 (37), RVSSG-8102 (39), RVSSG-40 (40)	ICCV 11110 (78), RVSSG-40 (78), AKG-9826 (79), RVSSG-8102 (79), NBeG 134 (82)	NBeG 139 (6.5), NBeG 169 (7), NBeG 379 (7.5)	NBeG 379 (45.4), ICC 9590 (46), RVSSG-8102 (46.7), NBeG 138 (46.9)	NBeG 803 (51.4), ICC 2942 (53.7), NBeG 134 (57.7), ICC 12479 (58.3), NBeG 141 (66.4)		NBeG 169 (9.8), RVSSG-40 (11.9)
X	49	NBeG 422 (42), NBeG 453 (43), ICC 1194 (43), ICC 6279 (43), B-6 (43)	ICC 4944 (80), B-6 (81), ICC 67 (82), ICC 4639 (82)	NBeG 639 (6.5), NBeG 380 (7), ICC 867 (7.5)		NBeG 143 (41), ICC 19360 (41.2), NBeG 453 (41.4), ICC 89220 (42.5), ICC 7441 (43.3)	ICCV 07112 (31.8), ICCV 08107 (31.9)	NBeG 639 (9.9), ICC 19049 (10.3), ICCV 07103 (11.1), ICCV 07112 (11.4), ICCV 08107 (11.9), NBeG 422 (12.1), B-6 (13)
XI	58	NBeG 171 (44), NBeG 382 (44), NBeG 452 (44)	NBeG 87 (80), ICC 5383 (82), ICCV 04105-1 (82), NBeG 677 (83), ICCV 09104 (83), ICCV 10939 (83)	ICCV 09101 (6), ICCV 09114 (6), ICCV 10939 (6), ICCV 09104(7)				NBeG 431 (11.5)
XII	66	NBeG 201 (44), GL-12021 (44), JG-130 (44)	NBeG 201 (82), BBG-2 (82), ICCV-5 (82), JG-36, (82) JG-130 (82)	NBeG 577 (6.5), ICCV 98004 (6.5), ICCV 98502 (7.5), ICCV 09107 (8)	ICC 95806 (48.5), ICC 12482 (48.7), ICC 18432 (49.7), BG-3062 (50), ICC 19176 (51.8)		ILWC-282CE (31.1), IPC-2012-30 (31.2), ICCV 16187 (33.3), JAC-29 (33.4), ICC 7074 (35)	PBC-1103 (10.5)
Total	**537**							

cluster IX recorded higher values for number of pods per plant and seed yield per plant. The genotypes of cluster VIII recorded low values for days to 50 per cent flowering and days to maturity. Therefore, the genotypes belonging to these clusters would be subjected to intercrossing (Balasaheb *et al.,* 2018 and Reddy *et al.,* 2021) and thus creating variability for respective characters and further improvement in seed yield. Considering the cluster distances and cluster mean value in the present investigation, emphasis should be given to genotypes belonging to clusters IV, VI, VIII and IX which could be recommended for further use in the breeding programme for the development of transgressive segregants for yield and yield component traits.

The data on inter cluster (**Table 2**) and *per se* performance of the genotypes (**Table 1**) were used to select genetically diverse and superior genotypes. The genotypes which were exceptionally good for one or more characters from these clusters would be more desirable for the development of transgressive segregants for yield attributing traits. Germplasm accessions with high *per se* performance for seed yield per plant from clusters IV, IX, VI and VII were NBeG-169, RVSSG-40, CP 117-5-23 and ICCV 07101, respectively. Early maturing genotypes which were genetically diverse were NBeG 829, NBeG 637, NBeG 625, ICCV 10107, AKG-9826, RVSSG-8102, RVSSG-40, NBeG 132, NBeG 138, NBeG 139, and ICC 5716. The genotypes *viz.,* CP-232-6-12, IPC-08-11, IPC-2010-134, CP-116-14-23, CP-117-5-1, CP-116-15-4-P4, CP-117-11-9, ICC 10945, CP-117-11-10 and CP-116-14-1 were the diverse genotypes with semi erect growth habit and could be utilised in breeding chickpeas suitable for combine harvesting. Development of chickpea cultivars with 30 to 40% more height than the existing cultivars and semi-erect to erect growth habit will make these cultivars suited to mechanical harvesting (Vishnu *et al*., 2018). For yield contributing characters like number

Table 2. Genetic diversity in 537 germplasm lines of chickpea as measured by intra and inter cluster distances between clusters

Cluster number	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
I	26.52	17.83	28.03	19.07	26.21	16.65	12.23	19.09	13.90	14.42	14.28
II		22.76	15.62	26.83	15.38	29.46	31.96	30.05	22.70	19.14	14.53
III			19.06	21.14	29.52	15.77	16.69	31.57	20.76	10.33	14.19
IV				19.91	29.90	21.88	27.31	**39.55**	29.64	18.69	14.71
V					34.17	14.69	14.87	35.12	29.25	21.29	14.50
VI						**36.52**	35.99	19.47	18.79	24.69	21.47
VII							11.05	34.27	24.64	14.45	16.37
VIII								31.12	24.51	18.29	18.18
IX									13.58	25.85	25.81
X										13.51	17.49
XI											10.41
XII											

Table 3. Cluster means of traits in different clusters as per K-means clustering and distribution of genotypes among 12 clusters

Cluster number	Days to 50% flowering	Days to maturity	Number of primary branches	Plant height (cm)	Number of pods	Seed yield per plant (g)	100 seed weight (g)
I	42.23	79.93	4.23	37.60	29.23	7.45	25.53
II	60.30	95.26	3.83	47.98	24.50	5.29	22.93
III	49.08	89.98	3.24	32.84	20.21	6.90	33.59
IV	56.31	92.89	3.89	44.64	10.50	3.15	26.20
V	41.65	80.30	3.91	47.59	13.57	4.21	28.20
VI	57.70	93.95	4.08	48.97	39.39	6.77	21.24
VII	41.62	81.80	3.85	33.33	13.71	3.74	25.02
VIII	40.16	76.28	3.76	35.39	19.78	6.20	31.52
IX	44.77	84.41	4.05	40.63	46.18	7.59	19.09
X	48.67	88.10	4.06	34.09	35.61	7.11	19.89
XI	49.41	89.50	3.73	33.45	22.90	5.43	23.78
XII	48.99	87.94	4.05	43.61	21.92	5.41	24.90

of primary branches diverse genotypes with high *per se* performance were CP-116-15-4-P2, ICCV 10113, ICCV 11102, ICCV 11106, ICCV 07104, NBeG 139, NBeG 169, and NBeG 379 where as the lines *viz.,* ICC 10939, CP-211-2-16, CP-116-14-17, CP 117-5-23, BDNG-2010-1, NBeG 803, ICC 2942, NBeG 134, ICC 12479 and NBeG 141 were promising for number of pods. These prospective genotypes could be utilized in hybridization programmes for evaluating their combining ability and to further endorse their utility in breeding *desi* chickpeas specifically adapted to Andhra Pradesh state.

REFERENCES

Annual report. 2020-21. All India Co-ordinated Research Project on Chickpea, IIPR, Kanpur, Uttar Pradesh.

Balasaheb, B.A., Akshays, S.M., Shridhar, B.G. and Suresh, B.G. 2018. Genetic diversity studies in chickpea (*Cicer arietinum L.*) germplasm. *International Journal of Current Microbiology and Applied Science,* **7**(9):2757-2763. [Cross Ref]

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

Gaur, P.M., Jukanti, A.K. and Varshney, R.K. 2012. Impact of genomic technologies on chickpea breeding strategies. *Journal of Agronomy*, **2**:199- 221. [Cross Ref]

Jayalakshmi, V. and Ronald, G.R. 2011. Assessment of genetic diversity for quantitative traits in chickpea (*Cicer arietinum* L.). *Journal of Research ANGRAU,* **39**:58-61.

Jayalakshmi, V., Kiran Kumar Reddy, C., Jyothirmayi, G. and Trivikram Reddy, A. 2016. Studies on genetic diversity in chickpea utilizing morphological and total protein markers. *Journal of Legume Research,* **39**:323-325. [Cross Ref]

Jayalakshmi, V., Ronald, R.G. and Lakshmanna, K. 2014. Diversity analysis of chickpea germplasm in scarce rainfall zone of Andhra Pradesh, India. *Journal of Legume Research,* **37**:682-684. [Cross Ref]

Jeethava, A.S., Yasufzal, A. S., Poshiya, V.K. and Vaddoria, M.A.2000. Divergence analysis in chickpea. *Gujrat Agricultural University Research Journal,* **22**(1):23-28.

Kuldeep, R., Pandey, S., Babbar, A. and Prakash, V. 2015. Genetic diversity analysis in Chickpea grown under heat stress conditions of Madhya Pradesh. *Electronic Journal of Plant Breeding*, **6**(4): 962-971.

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

Parameshwarappa, S.G., Salimath, P.M., Upadhyaya, H.D., Patil, S.S. and Kajjidoni, S.T. 2011. Genetic divergence under three environments in a minicore collection of chickpea (*Cicer arietinum* L.). *Indian Journal of Plant Genetic Resources,* **24**(2):177-185.

Parashi, V.S., Lad, D.B., Mahse, L.B., Kute, N.S. and Sonawane, C.J. 2013. Genetic diversity studies in chickpea (*Cicer arietinum* L.). *Bioinfolet,* **10**(1):337-341.

Raval, L.J. and Dobariya, K. 2004. Assessment of genetic divergence in chickpea (*Cicer arietinum* L.). *Annuals of Agricultural Research,* **25**(1):30-34.

Reddy, J.M., Lal, G.M., Reddy, V.P., Pattanayak, S., Guptha, V.R, Sagar, C.K. and Brahmanjaneyulu, V.P.B. 2021. Studies on genetic diversity in *Desi* chickpea (*Cicer arietinum* L.) using D^2 Statistics. *International Journal of Plant & Soil Science,* **33**(17):100-104. [Cross Ref]

Sreelakshmi, Ch., D. Shivani and C.V. Sameer Kumar. 2010. Genetic divergence, variability and character association studies in bengalgram (*Cicer arietinum* L.). *Electronic Journal of Plant Breeding*, **1**(5): 1339-1343.

Van der Maesen, L.J.G. 1972. Origin, history and taxonomy of chickpea. In: The chickpea (Eds. Saxena, M.C. and Singh, K.B.). C.A.B International and ICARDA, Wallingford, UK, 11-34.

Varshney, R.K., Song, C., Saxena, R.K., Azam, S., Yu, S. and Sharpe, A.G. 2013. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature Biotechnology,* **3**:240-246. [Cross Ref]

Vishnu, B., Jayalakshmi, V. and Sudha Rani, M. 2018. Genetic diversity studies among chickpea (*Cicer arietinum* L.) genotypes under rainfed and irrigated conditions for yield attributing and traits related to mechanical harvesting. *Legume Research,* DOI: 10.18805/LR-3959. [Cross Ref]

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