RESEARCH ARTICLE Agricultural Science Digest, Volume 43 Issue 5: 661-667 (October 2023)

Unraveling the Relationship between Fruit Yield and Yield Related Components in Snake Gourd Genotypes using Multivariate Analysis

A. Fahima Fathima[1], L. Pugalendhi[1], T. Saraswathi[2], N. Manivannan[3], M. Raveendran[4]

10.18805/ag.D-5753

ABSTRACT

Background: Snake gourd is a monoecious crop that prefers cross pollination. Snake gourd has a lot of potential for genetic improvement. A large variation can be produced when genetically diverse and geographically distant lines are combined. To examine the genetic diversity and connection between essential agronomic features in snake gourd, multivariate methods such as principal component analysis and cluster analysis were used. This study will use multivariate analysis to determine the genetic diversity and link between critical agronomic aspects of snake gourd.

Methods: A total of sixteen genotypes and two varieties of snake gourd genotypes were subjected to boxplot, principal component analysis and cluster analysis based on eleven quantitative traits. Boxplot analysis, Principal component analysis and cluster analysis were performed using R version of 4.2.1.

Result: Boxplot analysis depicted the frequency distribution of eleven quantitative traits among 18 snake gourd accessions. The overall variation was split into eleven principal components, out of which five major principal components contributed for variability of snake gourd genotypes by exhibiting 90.05 per cent of variability. The squared cosine variables inferred that the traits *viz*., days to first male flowering, days to first female flowering and days to first harvest contributed more for variability in the first component. The ward D2 method of hierarchical clustering cluster the 16 genotypes and 2 varieties in two clusters based on cluster sum of squares.

Key words: Boxplot, Cluster, Principal components, Squared cosine.

INTRODUCTION

Trichosanthes anguina L. (snake gourd) is a monoecious crop that prefers cross pollination. Because of its outcrossing features, variability is frequently generated. Genetic diversity is vital in any crop development programme. According to Ahmed *et al*. (2000), there is a lot of possibility for genetic improvement in snake gourd because the germplasm has a lot of variability. The nutritional value of green fruits is exceptional and they compare favorably with the nutritional value of any other type of vegetable. For every 100 g of edible fruit, the fruit has a respectable quantity of carbohydrates (3.3%), protein (0.5%), minerals (0.5%), fiber (0.5%) and fat (0.3%). Phosphorus (135 mg/100 g), potassium (121.6 mg/100 g), magnesium and zinc are the primary mineral elements that can be found in snake gourd. (Ojiako and Igwe, 2008). In general, snake gourd fetches a higher price per unit of land, but the average yield of the crop is low in India when compared to other neighbouring countries and its production is also restricted to only three to four months of the year (Kumar *et al*., 2022). When genetically varied and geographically distant lines are united, a wide range of variation can be produced. The inherent low yield potential of varieties, a lack of genetic variation and inefficient plant types, weed competition, a lack of appropriate plant ideotypes and vulnerability to biotic and abiotic stress are the key roadblocks to bolstering yields. Several academicians have studied about genetic diversity, clustering patterns and the relative contributions of various features to divergence and selection effectiveness. Given the aforementioned, genetic divergence information in these genotypes would aid in the selection of potential parental materials for a successful breeding programme. Crop genetic improvement, at the very least, consists of three distinct aspects: genetic variability, genotype selection and evaluation. The evaluation of genetic

---

[1]Department of Vegetable Science, Horticulture College and Research Institute, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India.
[2]Department of Medicinal and Aromatic Crops, Horticulture College and Research Institute, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India.
[3]Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India.
[4]Directorate of Research, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India.

Corresponding Author: A. Fahima Fathima, Department of Vegetable Science, Horticultural College and Research Institute, Tamil Nadu Agricultural University, Coimbatore-641 003, Tamil Nadu, India. Email: fahimarahuman@gmail.com

How to cite this article: Fathima, A.F., Pugalendhi, L., Saraswathi, T., Manivannan, N. and Raveendran, M. (2023). Unraveling the Relationship between Fruit Yield and Yield Related Components in Snake Gourd Genotypes using Multivariate Analysis. Agricultural Science Digest. 43(5): 661-667. doi: 10.18805/ag.D-5753.

Submitted: 06-03-2023 **Accepted:** 23-05-2023 **Online:** 10-06-2023

diversity is the most important step since it is crucial in determining the breeding strategies. Furthermore, selecting targeted attributes, particularly those that are not apparent to the naked eye, may be troublesome.

Multivariate analysis, a statistical technique encompassing several variables, has been utilized to evaluate the genetic variation of plants and establish the relationship between their traits. Multivariate analysis is a powerful tool to assist initial stage of crop improvement as it allows many traits to be evaluated simultaneously (Barth *et al*., 2022). Numerous crop species have demonstrated the efficacy of multivariate analysis for evaluating genetic diversity and trait correlations. In this study, multivariate methods such as principal component analysis and cluster analysis were employed to assess the genetic diversity and association between important agronomic characteristics in snake gourd. The study is carried out to assess the degree of genetic variation among genotypes and to identify different parents for future genetic studies of snake gourd.

MATERIALS AND METHODS

The current study was conducted at the Vegetable Research Fields, College Orchard, HC and RI, Coimbatore, during June to October 2021. Sixteen genotypes and two varieties of snake gourd (Table 1) were collected from various regions of Tamil Nadu and the experiment was conducted in a randomized block design with two replications. Each replication includes ten pits with two plants per pit. The spacing between pits and blocks was 2.0 m and 2.0 m, respectively. Multivariate analysis was performed for 18 accessions based on the characteristics of node order of male flower, node order of female flower, days to first male flowering, days to first female flowering, internodal length, days to first fruit harvest, fruit length, fruit girth, single fruit weight, number of fruits per plant and fruit yield per plant. Under multivariate analysis, R studio version 4.2.1 (R core team, 2021) was used to conduct box plot analysis, principal component analysis (PCA) and cluster analysis. For box plots, PCA and cluster analysis, the packages gplot, factoMineR and factoextra were used.

RESULTS AND DISCUSSION

Boxplot analysis

Boxplots were constructed for eleven quantitative traits to know the phenotypic variation between and within the eighteen germplasm collected from different geographic origins. A box and whisker plot is a graph that displays a visual representation of a data set's statistical five-number summary including sample minimum score, first (lower quartile), median and third (upper) quartile represents the 25th,50th and 75th percentile of the accessions respectively (Krishna *et al*.,2022). The frequency distribution for eleven quantitative traits across 18 snake gourd accessions were presented in the form of box plots to highlight genetic diversity (Table 2, Fig 1). The population variation is represented by the vertical lines (whiskers). Outliers are

Table 1: Treatment details and source.

Accessions	Treatment details	Source
TA01	Pallapatti	Karur
TA02	Vilathikulam short	Thoothukudi
TA03	Kariyapatty	Virudhunagar
TA04	Musiri (Long)	Trichy
TA05	Madurai (Long)	Madurai
TA06	Kamudhi	Ramanathapuram
TA07	Madurai (Short)	Madurai
TA08	Kalappal	Thiruvarur
TA09	Trichy Short	Trichy
TA10	Trichy Long	Trichy
TA11	Olakkur	Villupuram
TA12	Tanjore	Tanjore
TA13	Kaduvaguli	Kerala
TA14	Poondhurai	Erode
TA15	Sivagiri	Erode
TA16	Gobichettipalayam	Erode
	CO-2	TNAU, Coimbatore
	PKM- 1	HC and RI, TNAU, Periyakulam

Table 2: Frequency distribution of eleven quantitative characteristics among 18 genotypes.

Statistic	IL	FG	FL	DMF	DFF	NMF	NFF	DFH	SFW	NOF	FY
Number of observations	18	18	18	18	18	18	18	18	18	18	18
Minimum	11.3	7.60	22.7	24	35.9	3.5	10.40	47.9	169.6	5.5	2061
Maximum	18.55	18.70	170.38	56.6	67.8	9.4	18.50	78.3	892.7	17.6	7513
1st Quartile	13.28	12.61	29.53	33.12	40.05	4.425	13.20	51.27	313.5	7.3	2911
Median	15.72	13.83	43.66	34.4	42.55	5.3	14.45	52.75	435.7	7.6	3407
3rd Quartile	16.61	15.58	52.59	37.98	47.65	7.45	16.10	58.7	644.3	10.1	5204
Mean	15.1	14.06	52.52	36.36	45.06	5.822	14.53	56.12	481.5	9.456	4050
Variance (n-1)	4.805038	7.493635	1362.79	49.48958	66.69556	3.631242	4.558595	59.67088	55143.84	15.99438	3008065
Standard deviation	2.19204	2.73745	36.91598	7.034883	8.166735	1.905582	2.135087	7.724693	234.8273	3.999297	1734.377

IL- Internodal length, FG- Fruit girth, FL- Fruit length, DMF- Days to first male flowering, DFF- Days to first female flowering, NMF- Node order of first male flower, NFF- Node order of first female flower, DFH- Days to first fruit harvest, SFW- Single fruit weight, NOF- Number of fruits per plant, FY- Fruit yield per plant.

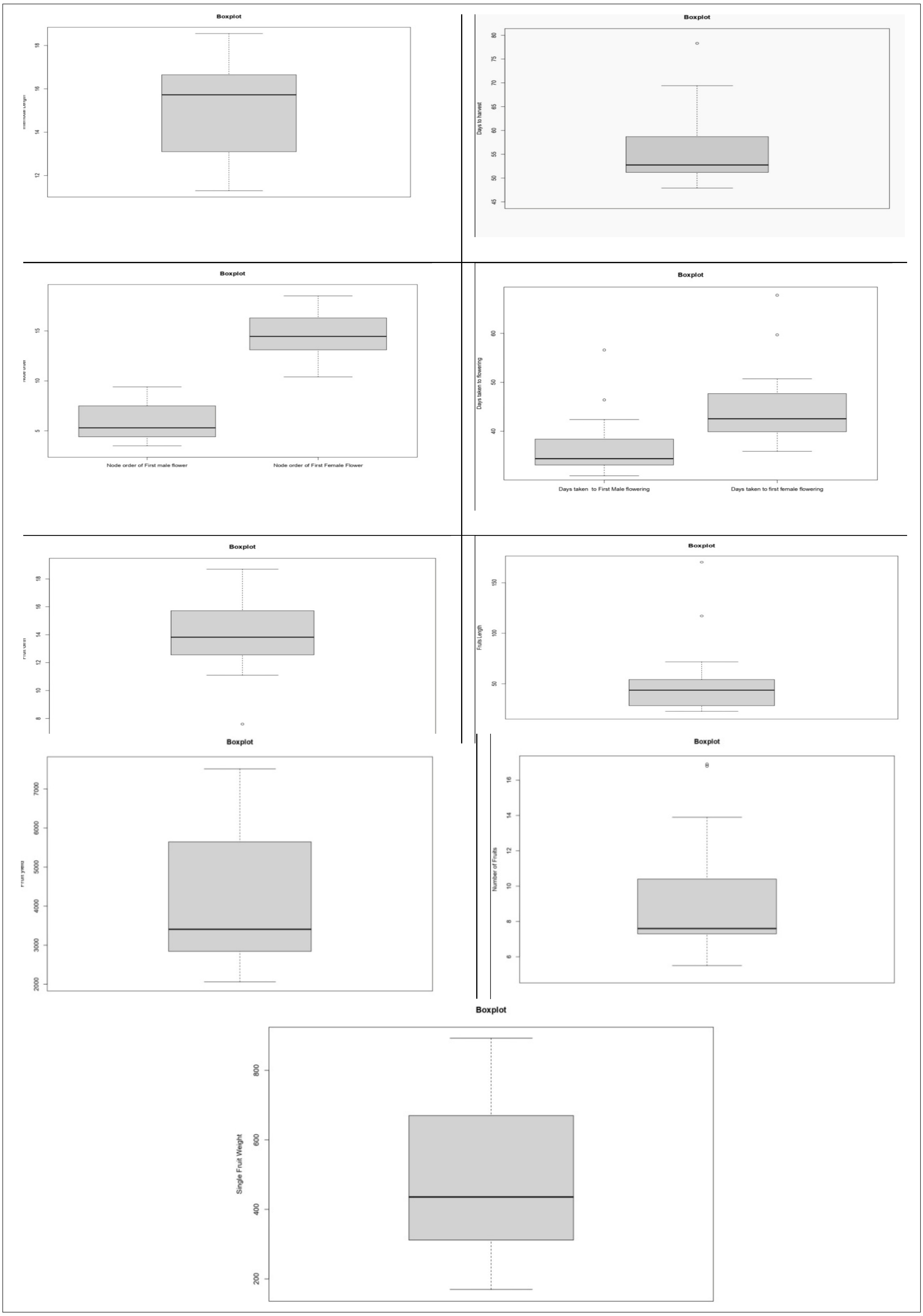

Fig 1: Boxplot depicting the variation in snake gourd genotypes.

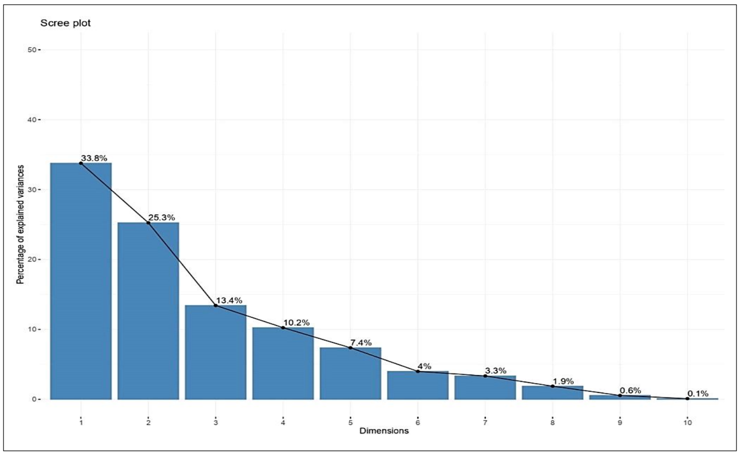

Fig 2: Scree plot for per cent variation of principal components based on yield attributing traits on snake gourd.

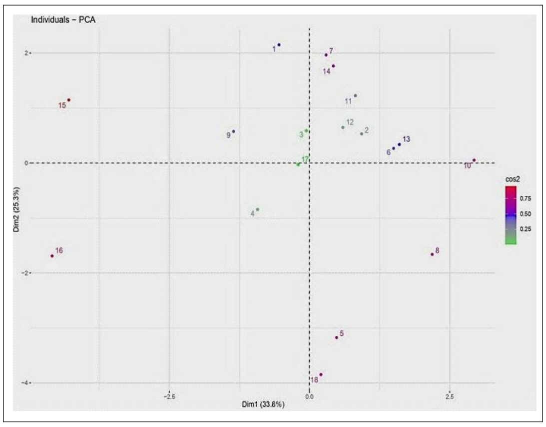

Fig 3: Contribution of each snake gourd genotypes towards the cumulative variability.

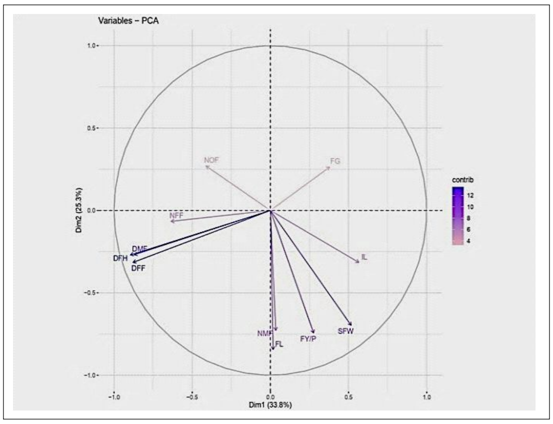

Fig 4: Contribution of each variables towards the variability in first two principal components.

drawn with dots. Similar results were also reported by Lakshmi *et al.* (2019) and Krishna *et al.* (2022) in rice.

PCA analysis

Principal Component Analysis is a well-known dimension reduction method that may be used to reduce a large set of interrelated variables to a small set that contains the majority of the information available in the large set (Singh *et al.*, 2020). The result of the principal component Analysis has shown the genetic diversity of the snake gourd germplasm lines.

Using a scree plot, the proportion of variance related with eigenvalues and principal components was depicted for each principal component (PC) derived from a graph. PC 1 demonstrated the greatest variation, 33.76 per cent, with an eigenvalue of 3.71, which steadily decreased in the remaining principal components. Up until the third Principal component, a semi-curved line was noticed; beyond that, a straight line with minimal variation in each Principal Component was detected. PC1 through PC11 have eigen values ranging from 3.71 to 0.009 According to Kaiser's criterion (eigen value >1), conditions are regarded satisfied (Kaiser, 1958). PC1 explains 33% of the variance, followed by PC2 (25.25%), PC3 (13.41%) and PC4 (10.24%). Table 3 revealed that, out of eleven main components, five displayed >0.5 eigenvalues and approximately 90.05 per cent variability, whereas four exhibited >1 eigenvalues and around 82.68 per cent variability. The variability among genotypes using principal component analysis has been reported by Karunakar *et al.* (2022) in moringa. It is evident

Table 3: Eigen value, variance per cent and cumulative variance component for the principal component analysis.

Principal components	Eigen value	Variance per cent	Cumulative variance per cent
PC1	3.714495256	33.77	33.77
PC2	2.777736962	25.25	59.02
PC3	1.476157084	13.42	72.44
PC4	1.127075314	10.25	82.69
PC5	0.810395746	7.37	90.05
PC6	0.439272349	4.00	94.05
PC7	0.36785355	3.34	97.39
PC8	0.206241936	1.88	99.27
PC9	0.060677593	0.55	99.82
PC10	0.010980295	0.09	99.92
PC11	0.009113916	0.08	100

from the graph (Fig 2) that PC1 exhibited the greatest variation compared to the other eleven PCs; consequently, the selection of lines for characters under PC1 may be desired. Verma *et al*. (2017) also reported a similar curve line in pointed gourd accessions.

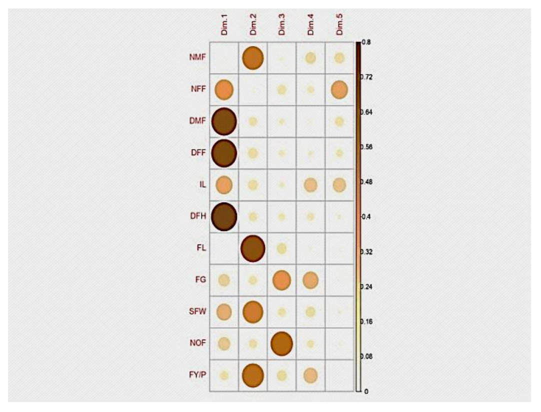

Fig 5: The contribution of measured characteristics on extracted principal components based on square cosine values.

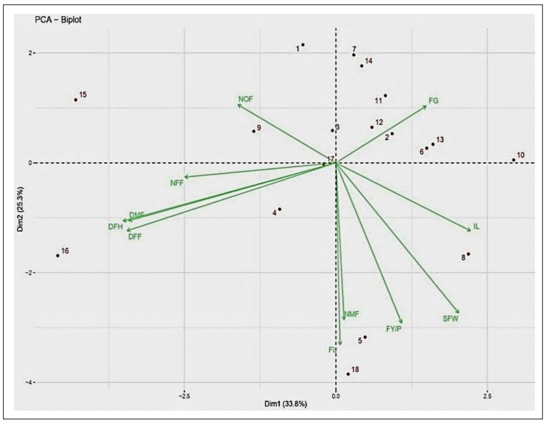

Fig 6: Biplot on basis of PC1 and PC2 for genotypes and variables.

The first four PCs in the rotated component matrix showed to have the highest variability (82.68%), suggesting that traits falling within these PCs may be prioritised in snake gourd breeding. It was shown that the internodal length, fruit girth and single fruit weight were the main factors contributing to the first principal component (PC1), which explained the majority of variation. Therefore, PC 1 permits simultaneous selection of that particular phenological trait, but other PCs permit simultaneous selection of other related traits. Each component's PC scores included both positive and negative values (Table 4). These scores could be used to construct precise selection indices based on the variability that each primary component can explain (Rahevar *et al*., 2021). PC1, PC2 and PC4 contained the majority of the yields contributing traits, according to PCA. Fig 3 and Fig 4 shows how each trait and variable contributed to overall variability in the genotyping studies. The results from the present study are equivalent to those from Belay *et al.* (2019) and Rahevar *et al.* (2021). The squared cosine variables based on eleven quantitative characters and five major principal components are depicted in Fig 5 which infers that high cosine square values for traits *viz*., days to first male flowering, days to first female flowering, days to first harvest in fruit component indicates good representation of such traits in the first component. This is in accordance with the results of Vijayakumar *et al.* (2020) who have explained the squared cosine variables on five major principal components in Indian cowpea.

To ascertain how the evaluated genotypes and characteristics were categorised, biplot analysis was utilised preceding PCA (Fig 6). Commercial varieties CO_2 and PKM1 were discovered to cluster with other genotypes using the biplot technique. The results of correlation analysis may be validated using biplot analysis, which also revealed the relationships between the characters. The biplot's narrow angle between the qualities reveals a high correlation between them (Cai *et al*., 2018). Days to first female flowering, days to first male flowering, days to first fruit harvest and number of fruits per plant were exhibited the oppositional narrow

Table 4: Contribution of variables for the principal component analysis.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
NMF	0.031148	19.12363	0.520408	11.05912	13.62539	38.70289	14.53225	3.72E-01	0.912343	0.05061	1.070087
NFF	10.86767	0.158109	5.63444	4.08876	40.83458	2.206576	33.08492	9.03E-04	0.126997	0.097429	2.899618
DMF	20.42641	2.607184	1.926842	0.413649	10.64974	2.813368	0.009022	6.83E-02	60.76157	0.01383	0.310048
DFF	20.78714	3.587697	2.207433	1.871455	6.67444	0.006307	0.820281	7.95E-02	20.83989	35.03184	8.093996
IL	8.625417	3.59209	1.675897	17.84757	24.62113	5.609038	34.73188	6.98E-02	3.040686	0.002626	0.183836
DFH	21.56086	2.631607	3.178147	4.598709	1.698631	0.001669	0.241111	8.72E-02	11.11457	37.08145	17.80605
FL	0.008076	25.71216	7.091824	0.316501	0.018493	18.77677	0.006185	4.57E+01	0.380195	0.298912	1.738462
FG	3.877587	2.496238	26.55969	25.30566	0.198374	9.464661	1.201346	2.98E+01	0.465868	0.651275	0.01131
SFW	7.181949	17.51156	4.650328	9.174032	1.017731	3.521802	0.924816	2.09E+01	2.32416	10.672	22.1547
NOF	4.571721	2.62685	39.29305	5.076839	0.564912	16.1055	11.35648	4.40E-01	0.016367	7.141566	12.80633
FY/P	2.062018	19.95288	7.261948	20.24771	0.096579	2.791415	3.091712	2.59E+00	0.017352	8.958455	32.92556

IL- Internodal length, FG- Fruit girth, FL- Fruit length, DMF- Days to first male flowering, DFF- Days to first female flowering, NMF- Node order of first male flower, NFF- Node order of first female flower, DFH- Days to first fruit harvest, SFW- Single fruit weight, NOF- Number of fruits per plant, FY- Fruit yield per plant.

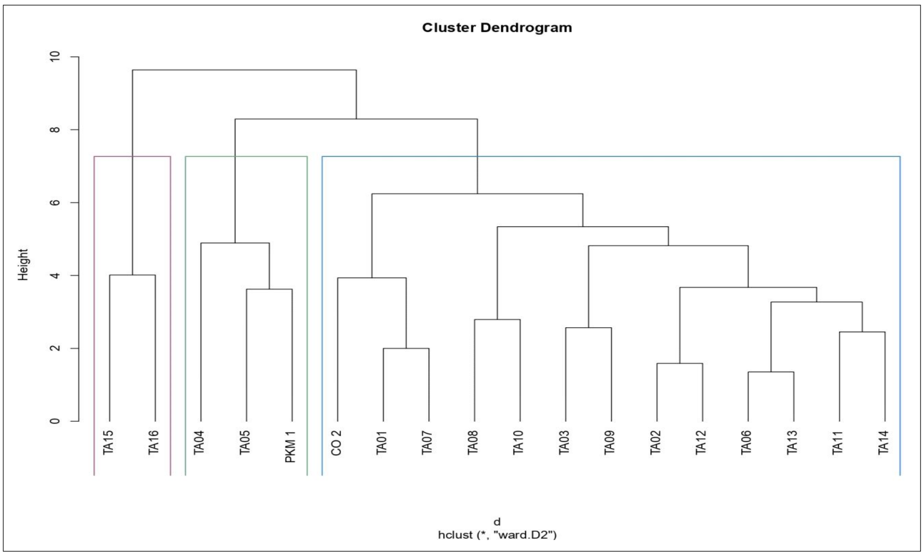

Fig 7: Hierarchical cluster dendrogram based on ward. D2 method showing the phenotypic relationship among snake gourd genotypes.

Table 5: Cluster means.

Cluster	NMF	NFF	DMF	DFF	IL	DFH	FL	FG	SFW	NOF	FY/P
1	-0.14286	1.719941	2.151975	2.289097	-1.08118	2.295668	-0.26903	-0.39595	-0.66367	0.661227	-0.50076
2	0.017857	-0.21499	-0.269	0.28614	0.135148	0.28696	0.033628	0.049494	0.082959	-0.08265	0.062595

angular direction of traits, which suggested a strong negative association between them. Widyavan *et al.* (2020) presented a similar biplot analysis in the Yardlong bean that explained the correlation between traits.

Cluster analysis

Woodyard's Hammock method or wards method of clustering has been used to do cluster analysis of snake gourd genotypes among the traits. Wards method of clustering uses the agglomerative clustering algorithm which forms cluster based on analysis of variance instead of distance matrix or measures of association. The cluster sum of squares was 24.8% which in turn indicates variability of traits within a cluster. The cluster means of traits for the 2 cluster have been presented in Table 5 in which days taken for first female flowering and days taken to first fruit harvest exhibited maximum cluster means in cluster 1 and cluster 2. Cluster analysis grouped 16 genotypes and 2 varieties into two clusters based on traits as shown in Fig 7. Cluster I comprised of genotypes TA15 and TA16 followed by genotypes TA04, TA05, PKM1, TA07, TA08, TA10, TA03, TA09, TA02, TA12, TA06, TA13, TA11, TA14 in cluster II. Similar Ward D2 method of clustering has been used for genetic relationships among yam accessions by Agre *et al.* (2019).

CONCLUSION

The frequency distribution of 18 snake gourd accession displayed the data sets statistical five set summary *viz.*, sample minimum score, first (lower quartile), median and third (upper) quartile represents the 25th,50th and 75th percentile of the accessions respectively. The results of principal component analysis showed that first four principal components contributed to the highest variability. Ward D2 method of clustering clustered the 18 genotypes into two clusters. Cluster II comprised of maximum number of genotypes followed by cluster I. The cluster mean sum of squares was 24.8 %.

Conflict of interest: None.

REFERENCES

Agre, P., Asibe, F., Darkwa, K., Edemodu, A., Bauchet, G., Asiedu, R., Adebola, P., Asfaw, A. (2019). Phenotypic and molecular assessment of genetic structure and diversity in a panel of winged yam (*Dioscorea alata*) clones and cultivars. Scientific Reports. 9(1): 18221. DOI: 10.1038/s41598-019-54761-3.

Ahmed, M.S., Rasul, M.G., Bashar, M.K. and Mian, A.S.M. (2000). Variability and heterosis in snake gourd. (*Trichosanthes anguina* L.). Bangladesh Journal of Plant Breeding and Genetics. 13(1): 27-32.

Barth, E., de Resende, J.T.V., Mariguele, K.H. (2022). Multivariate analysis methods improve the selection of strawberry genotypes with low cold requirement. Scientific Reports. 12: 11458. DOI: 10.1038/s41598-022-15688-4.

Belay, F., Abate, B. and T sehaye, Y. (2019). Genetic diversity studies for morphological traits of hot pepper (*Capsicum annuum* L.) genotypes in Central Zone of Tigray Region, Northern Ethiopia. African Journal of Agricultural Research. 14(33): 1674-1684. DOI: 10.5897/AJAR2019.14256.

Cai, T., Meng, X., Liu, D., Liu, T., Wang, H., Jia, Z., Yang, D., Ren, X. (2018). Exogenous hormonal application regulates the occurrence of wheat tillers by changing endogenous hormones. Frontiers in Plant Science. 9: 1-17. DOI: 10.3389/fpls.2018.01886.

Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika. 23: 187-200.

Karunakar, J., Preethi, T.L., Boopathi, M.N., Pugalendhi, L., Hepziba, S.J., Prasad, Y.M.M. (2022). Multivariate analysis through principal components for yield-attributing trials in indigenous moringa (*Moringa oleifera* L.) germplasm lines. International Journal of Plant and Soil Science. 34(9): 87-94. DOI: 10.9734/ijpss/2022/v34i930916.

Krishna, K., Mohan, Y.C., Shankar, V.G. and Rani, C.H. and Krishna, L. (2022). Principal component analysis and assessment of out cross potential traits associated with grain yield in maintainer lines of rice (*Oryza sativa* L.). Journal of Crop and Weed. 18. DOI: 10.22271/09746315.2022.v18.il.1549.

Kumar, M.J., Rekha, G.K., Ramana, C.V., Rajani, A. and Suneetha, D.R.S. (2022). Performance of snake gourd (*Trichosanthes anguina* L.) genotypes for growth, yield and quality traits in coastal Andhra Pradesh. The Pharma Innovation. 11(8): 232-235.

Lakshmi, V.I., Sreedhar, M., Vanisri, S., Anantha, M.S., Rao, L.S. and Gireesh. (2019). Multivariate analysis and selection criteria for identification of African rice (*Oryza glaberrima*) for genetic improvement of indica rice cultivars. Plant Genetic Resources. 17(6): 499-505. DOI: 10.1017/S1479262119000327.

Ojiako, O.A. and Igwe, C.U. (2008). The nutritive, anti-nutritive and hepatotoxic properties of (*Trichosanthes anguina* L.) (Snake tomato) fruits from Nigeria. Pakistan Journal of Nutrition. 7(1): 85-89. DOI:10.3923/pjn.2008.85.89.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.https://www.R-project.org.

Rahevar, P.M., Patel, J.N., Axatjoshi, Sushilkumar, Gediya, L.N. (2021). Genetic diversity study in chilli (*Capsicum annum* L.) using multivariate approaches. Electronic Journal of Plant Breeding. 12(2): 314-324. DOI: 10.379922/2021.1202.047.

Singh, P., Jain, P.K., Tiwari, A. (2020). Principal component analysis approach for yield attributing traits in chilli (*Capsicum annum* L.) genotypes. Chemical Science Review and Letters. 9: 87-91. DOI: 10.37273/chesci.cs232050121.

Verma, P., Maurya, S.K., Yadav, H., Panchbhaiya, A. (2017). Determination of genetic divergence in pointed gourd by principal component and non-hierarchical euclidean cluster analysis. Journal of Applied and Natural Science. 9: 2421-2426.DOI:10.31018/jans.v9i4.1548.

Vijayakumar, E., Thangaraj, K., Kalaimagal, T., Vanniarajan, C., Senthil, N., Jeyakumar, P. (2020). Multivariate analysis of 102 Indian cowpea [*Vigna unguiculata* (L.) Walp.] germplasm. Electronic Journal of Plant Breeding. 11(1): 176-183. DOI: 10.37992/2020.1101.031.

Widyawan, M.B., Hasanah, A., Taryono, Alam, T., Sayekti, R.S., Praman, A.A.C., Wulandari, R.A. (2020). Multivariate analysis unravels genetic diversity and relationship between agronomic traits, protein and dietary fiber in yardlong bean (*Vigna unguiculata* subsp, *sesquipedalis* Verdc.). Biodiversitas. 21. DOI: 10.13057/biodiv/d211211.