

Principal component analysis, a multivariate technique to assess yield-contributory traits in the sesame Germplasm

D. Umamaheswari¹, S.*Suganthi², S. Thirumeni³, P. Satheesh Kumar⁴ and R. Bhuvaneswari⁵

¹Department of Genetics and Plant Breeding, Faculty of Agriculture, Annamalai University, Annamalai Nagar, District-Chidambaram 608 002, T.N., India

²Department of Genetics and Plant Breeding, Agricultural College and Research Institute (AC & RI), Vazhavachanur, Tamil Nadu Agriculture University, Thiruvannamalai, India

³Department of Plant Breeding and Genetics, Pandit Jawaharlal Nehru College of Agriculture and research institute, Affiliated to Pondicherry University, Karaikal, India

⁴Department of Genetics and Plant Breeding, Anbil Dharmalingam Agricultural College and Research Institute (ADAC&RI), Tamil Nadu Agriculture University, Trichy, India

⁵Department of Soil Science and Agricultural Chemistry, Faculty of Agriculture, Annamalai University, Annamalai Nagar, District- Chidambaram 608 002, T.N., India

(Received 7 August, 2024; Accepted 12 October, 2024)

ABSTRACT

This study uncovers the genetic diversity of 40 sesame (*Sesamum indicum* L.) genotypes and identifies the key yield-contributory traits using Principal Component Analysis (PCA), a multivariate technique. Sesame is an important oilseed crop grown for seed oil and harbors greater genetic diversity besides it faces challenges like low productivity, vulnerable to biotic and abiotic stresses. PCA is commonly used in plant breeding for assessing the genetic diversity. The analysis revealed three principal components such as days to fifty percent flowering, number of primary branches per plant and number of capsules per plant accounting for 69% of the total phenotypic variation and its relationship with other components. PC1 explained the largest proportion of variation (41.99%), followed by PC2 (15.18%) and PC3 (11.79%). The Scree plot illustrates each component's contribution to phenotypic variation, while the biplot graph visualizes the genotype distribution. Thus, these principal components may serve as selection criteria for improving seed yield in sesame.

Key words: Sesame, PCA, Variation, Biplot

Introduction

Sesame (*Sesamum indicum* L.) is the most versatile oilseed crop, cultivated globally for its nutrition rich seed oil and protein. It is native to India and Africa and globally cultivated in many tropical and sub-

tropical regions for its edible oil, food, and fodder. India is the second-largest producer of sesame, accounting for approximately 40% of global seed production, FAOSTAT (2022). Sesame's importance extends beyond its economic value, as it also contributes significantly to food security, rural livelihoods,

(* Assistant Professor)

and environmental sustainability. However, sesame production faces challenges including low yields, productivity and vulnerability to climate change. To address these challenges encountered by sesame breeders' characterization of genetic variation in sesame accessions is crucial. The existence of genetic diversity within the germplasm offers a valuable resource for selecting diverse parents with desirable traits (Doijad *et al.*, 2016). By leveraging this diversity, breeders can exploit the segregation of genes in progeny, thereby increase the efficiency of selection process in breeding programs. Multivariate analysis is a statistically significant tool in analyzing the total variability of the multiple variables. It uses linear combinations of correlated variables to reduce the dimensionality of the data and estimates the contribution of variation created by each trait concurrently towards the principal components. Through multivariate analysis, researchers can evaluate the degree of population divergence and identify the major traits accounting for more observed variation. Sanwal *et al.* (2015). Principal component analysis (PCA) is a common multivariate technique used for data set transformation and reduction Mukhthambica *et al.* (2023). PCA analysis was widely used for studying genetic diversity in sesame by Gimbichu *et al.* (2023); Baraki *et al.* (2020) and Priyadharshini *et al.* (2018). Hence the present study focused on studying the diversity and variation exists on sesame germplasms by Principal component analysis. PCA reduces a large set of correlated variables into a smaller set of uncorrelated linear variables. These linear variables, called principal components, serve as crucial factors in selection process Hemant Kumar Singh and Bhim Singh (2013) thereby makes the plant breeder to understand and explore the significant traits and its contribution towards yield.

Materials and Methods

The present investigation was carried out in the Department of Plant Breeding and Genetics at

Pandit Jawaharlal Nehru College of Agriculture and Research Institute, Karaikal, during 2022. The experimental material consists of 40 genotypes and it was conducted in randomized block design with three replications. Nearly ten Quantitative traits such as days to fifty percent flowering (DFF), number of primary branches per plant (NPB), number of capsules per plant (NCP), Plant height (PH) (cm), days to maturity (DM), number of seeds per capsule (NSC), capsule length (CL) (cm), Capsule Breadth (cm), 1000 seed weight (TSW) (g) and Single plant yield (SYP) (g) were recorded. Principal Component Analysis was performed to identify and to estimate the relative contribution of various characters for total variability. Statistical analysis was done using Statistical Tool for Agricultural Research (STAR) package of IRRI, version 2.0.1, (2014).

Results and Discussion

The PCA scores of 40 sesame genotypes, with its ten principal components were calculated and depicted as PC1, PC2, PC3, PC4, PC5, PC6, PC7, PC8, PC9 and PC10. Under each principal component category the eigen values which recorded more than 1 (Table 1) were selected as main principal components Brejda *et al.* (2000) which mean that the particular trait accounted with minimum 10% of phenotypic variation. Among the ten components studied, totally three main principal components were elucidated in this investigation. They are Days to fifty percent flowering, number of primary branches per plant and number of capsules per plant. The fourth component plant height recorded the eigen value (0.989) which was very close to 1. These three components accounted to 69% of total observed phenotypic variation approximately. The proportion of variance of PC1, PC2, PC3 and PC4 was found to be 0.4199, 0.1518, 0.1179, 0.1090 respectively. PC1 (Days to fifty percent flowering) registered the highest portion of the variation accounting to 41.99% followed by number of primary branches per plant (15.18%) and number of capsules per plant with

Table 1. Principal component analysis for quantitative traits

Statistics	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2.0492	1.2322	1.0857	0.9949	0.8481	0.8033	0.6036	0.414	0.3829	0.2585
Proportion of Variance	0.4199	0.1518	0.1179	0.099	0.0719	0.0645	0.0364	0.0171	0.0147	0.0067
Cumulative Proportion	0.4199	0.5717	0.6896	0.7886	0.8605	0.9251	0.9615	0.9787	0.9933	1.0000
EigenValues	4.1993	1.5183	1.1788	0.9899	0.7192	0.6454	0.3644	0.1714	0.1466	0.0668

11.79%. In PCA usually the first component contributes for more variation subsequently the other components. The relationship of each trait studied with principal components gives clear representation about its interaction towards yield (Table 2). PC1 had positive values for thousand seed weight, single plant yield, plant height, capsule length, capsule breadth, number of capsules per plant and number of primary branches per plant. Similar positive interaction of PC1 for number of capsules per plant and single plant yield (g) was recorded by Sasipriya, *et al.* (2023) and Iqbal *et al.* (2016). The second principal component showed significant positive interaction with number of capsules per plant and capsule length whereas high negative interaction with days to maturity was obtained as declared by Bhagyashree *et al.* (2022), number of seeds per capsule and number of primary branches per plant. PC3 registered positive values with days to fifty percent flowering, single plant yield and thousand weight this was on contrary with Gupta *et al.* (2021). It also had high negative values with number of primary branches per plant and number of capsules per plant. Using PCA, a scree plot diagram (Figure 1) was portrayed between the principal components and eigen values taken on x and y ordinates which explains the total variability available in the germplasm Ezhilarasi *et al.* (2024). The principal component 1 with Eigen Values of 4.1993, PC2 with 1.5183 and PC3 with 1.1788 were documented. The succeeding principal components follow the lowest value which has less role in contribution towards the variation. The distribution of forty sesame germplasm using three principal components such

as PC 1, PC2 and PC 3 are plotted as biplot graph (Fig. 2-4). These biplot graph depicts how the genotypes were scattered throughout the graph in different groups, which undoubtedly exhibited the genetic diversity prevailing among the sesame gene pool. The length of the vector of each character in the biplot shows the variability in Fig.1, 2 and 3. In Fig.2 days to fifty percent flowering, days to maturity, thousand seed weight, seed yield per plant shows more variability. In Fig.3 single plant yield, thousand seed weight and capsule breadth exhibited greater variability when compared to other traits. In Fig. 4 days to fifty percent flowering, and

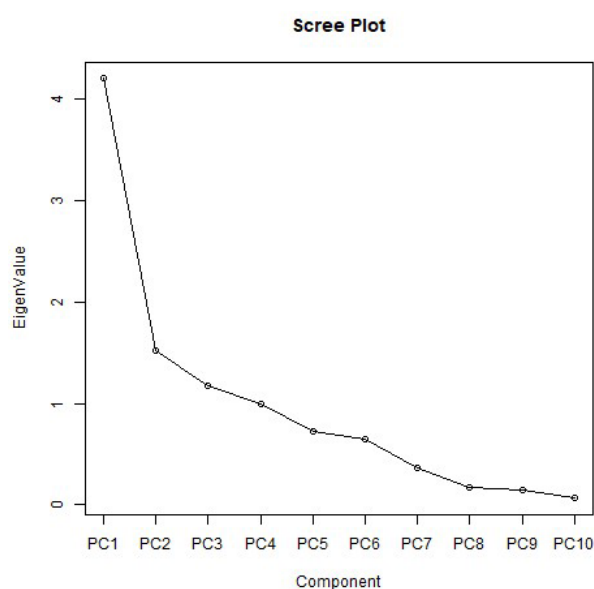


Fig. 1. Scree plot between principal components and Eigen values

Table 2. Factor loading by various traits

Variables	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
DFP	-0.0824	0.1775	0.4834	-0.7157	0.4347	-0.0463	0.0341	-0.0118	0.1498	-0.0294
NPB	-0.2011	-0.4807	-0.4276	-0.2699	0.2459	0.0821	-0.6232	0.1076	-0.0033	-0.0749
NCP	0.2295	0.3040	-0.2364	0.3434	0.7521	-0.2618	-0.0372	0.1012	-0.0411	0.1833
PHT	0.4212	-0.2453	0.1328	0.0003	-0.0434	0.2707	0.0304	0.5778	0.4215	0.3965
DM	-0.3094	-0.3679	-0.2761	-0.0319	0.3009	0.2613	0.7250	-0.0624	0.0481	0.0322
NSC	0.2303	-0.4570	-0.0132	-0.1253	-0.0983	-0.8094	0.1685	-0.1160	0.1110	0.0572
CL	0.2647	0.2701	-0.4750	-0.5085	-0.2092	0.0277	0.1376	0.0390	-0.4108	0.3729
CB	0.4029	0.1989	-0.3763	-0.1415	-0.0301	0.0948	0.0849	-0.1229	0.4944	-0.5999
TSW	0.4114	-0.2321	0.1386	0.0400	0.1313	0.3195	-0.1321	-0.7392	-0.0123	0.2695
SPY	0.4150	-0.2726	0.2171	0.0045	0.1448	0.1182	0.1022	0.2526	0.6084	-0.4763

DFP - days to fifty percent flowering; NPB - number of primary branches per plant; NCP - number of capsules per plant; PHT - Plant height(cm); D- days to maturity; NSC- number of seeds per capsule; CL-capsule length – cm; CB-capsule Breadth(cm); TSW – Thousand seed weight (g); SPY – single plant yield

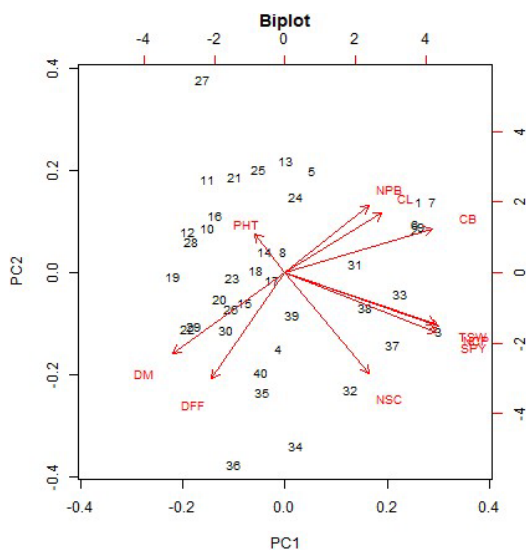


Fig. 2. Biplot Depicting Genotype dispersion between PC1 and PC2

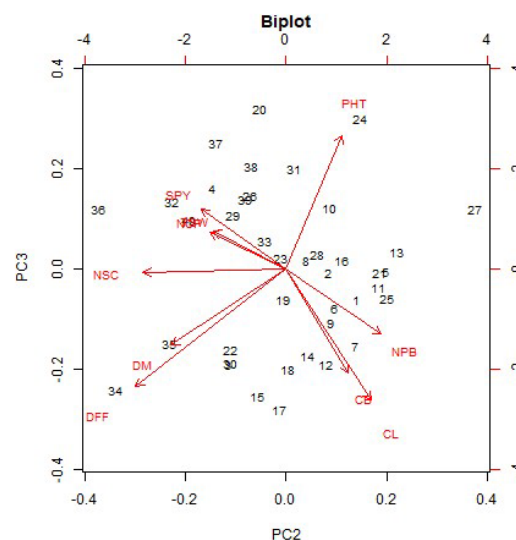


Fig. 4. Biplot Depicting Genotype dispersion between PC2 and PC3

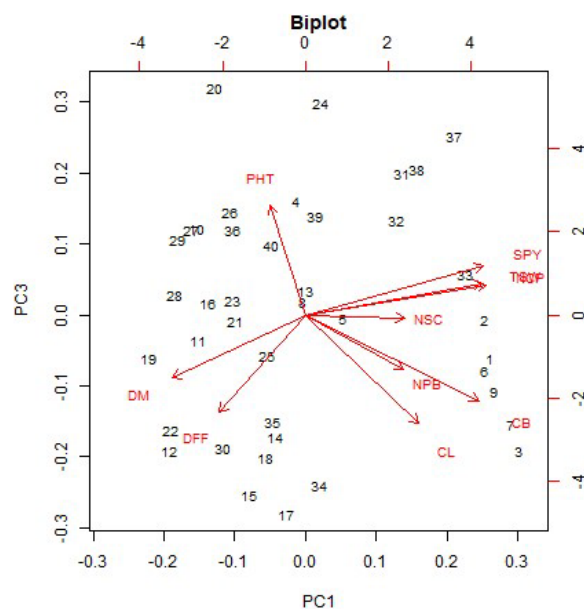


Fig. 3. Biplot Depicting Genotype dispersion between PC1 and PC3

plant height displayed the higher variability as reported by Ezhilarasi *et al.* (2024). Similar pattern of Principal component analysis and biplot graph was obtained by Karadi *et al.* (2019) in sorghum germplasm. Further in groundnut Sukrutha *et al.* (2023) attempted PCA and depict the biplot graph to elucidate the dispersal of genotypes across first two principal components for studying variation and diversity.

Conclusion

This study effectively utilized Principal component analysis (PCA) to assess genetic diversity and trait variability among 40 sesame genotypes. The PCA revealed three main principal components that accounted for approximately 69% of the total phenotypic variation. These findings underscore the importance of the three traits *viz.*, Days to fifty percent flowering, number of primary branches per plant and number of capsules per plant are choice of selection criteria in sesame breeding program aiming higher yield. The biplot graphs provided a clear depiction of the distribution of genotypes, highlighting the significance of genetic diversity within the sesame germplasm. Such exploration and understanding are highly rewarded in plant breeding program to select promising traits and parental lines for future crop improvement.

Conflict of interest

The author declares that there is no conflict of interest in publication of this paper.

References

- Baraki, F., Gebregergis, Z., Belay, Y., Berhe, M., Teame, G., Hassen, M., Gebremedhin, Z., Abadi, A., Negash, W., Atsbeha, A. and Araya, G. 2020. Multivariate analysis for yield and yield-related traits of sesame (*Sesamum indicum* L.) genotypes. *Heliyon*, 6(10).

- <https://doi.org/10.1016/j.heliyon.2020.e05295>
- Bhagyashree, D. D., Geethanjali, S. and Sasikala, R. 2022. Assessment of genetic variability for seed yield and its components in sesame (*Sesamum indicum* L.) based on multivariate analysis. *Electron. J. Plant Breed.* 13: 974-982 <https://doi.org/10.37992/2022.1303.126>
- Brejda, J. J., Moorman, T. B., Karlen, D. L. and Dao, T. H. 2000. Identification of regional soil quality factors and indicators I. Central and Southern High-Plains. *Soil Sci. Soc. Am. J.* 64: 2115-2124
- Doijad, S.B., Bagade, A.B. and More, A.W. 2016. Evaluation of sorghum germplasm for genetic diversity using D² statistics. *Electron. J. Plant Breed.* 7: 934-938
- Ezhilarasi, T., Mahalingam, A. and Manivannan, N. 2024. Assessment of genetic diversity in Indian sesame (*Sesamum indicum* L.) germplasm. *Eco. Env. and Cons.* 30: S344-S348. <https://doi.org/10.53550/EEC.2024v30I03S.059>
- FAOSTAT. Production - Crops and livestock products. 2022. <https://www.fao.org/faostat/en/#compare>
- Gimbichu, M.T., Gurmu, G.N. and Ayana, N.G. 2022. Genetic divergence analysis among sesame (*Sesamum indicum* L.) accessions in Middle Awash, Ethiopia. *Biorxiv.* <https://doi.org/10.1101/2022.12.29.522196>
- Gupta, D., Muralia, S., Khandelwal, V. and Nehra, A. 2021. Assessing diversity of sesame genotypes using cluster analysis and principal component analysis. *Int. J. Curr. Microbiol. App. Sci.* 10: 304-312
- Iqbal, A., Ahktar, R., Begum, T. and Dasgupta, T. 2016. Genetic estimates and diversity study in sesame (*Sesamum indicum* L.). *IOSR J. Agric. Vet. Sci.* 9: 1-5
- Karadi, A. and Kajjidoni, S.T. 2019. Genetic diversity and principal component analysis for grain quality and productivity traits in sorghum (*Sorghum bicolor* L. Moench). *Int. J. Curr. Microbiol. App. Sci.* 8: 839-850
- Mukthambica, K., Bisen, R. and Ramya, K.T. 2023. Principal component analysis for yield and yield-related traits in sesame (*Sesamum indicum* L.). *Biol. Forum Int. J.* 15: 227-232
- Priyadharshini, B., Prakash, M., Vignesh, M., Murugan, S. and Anandan, R. 2018. Genetic diversity analysis in sesame (*Sesamum indicum* L.) genotypes using principal component analysis. *Indian J. Agric. Res.* 52: 708-711 <https://doi.org/10.18805/IJAR.A-5080>
- Sanwal, S., Singh, B., Singh, V. and Mann, A. 2015. Multivariate analysis and its implication in breeding of desired plant type in garden pea (*Pisum sativum*). *Indian J. Agric. Sci.* 85: 1298-130.
- Sasipriya, S., Parimala, K., Eswari, K.B. and Balram, M. 2023. Principal component analysis and metroglyph of variation among sesame (*Sesamum indicum* L.) genotypes. *Electron. J. Plant Breed.* 13: 1234-1242. <https://doi.org/10.37992/2022.1304.156>
- Singh, H.K. and Singh, B. 2013. A note on multivariate technique in agricultural experiments. *Int. J. Curr. Sci.* 5: 50-56.
- Sukrutha, B., Kiran Kumar Reddy, C., Naga Madhuri, K.V., Bhargava Rami Reddy, Ch., Nirmal Kumar, A.R., Vemireddy, L.N. and Akkareddy, S. 2023. Principal Component Analysis and Path Coefficient Analysis for Groundnut Yield and Seed Quality Attributes (*Arachis hypogaea* L.). *Legume Res. an Int. J.* (1): 1-7. <https://doi.org/10.18805/LR-5075>