



# Correlation and Principal Component Analysis of Yield-related Traits in Castor (*Ricinus communis* L.) Genotypes

K. Chandana <sup>a\*</sup> and A.V.S. Durga Prasad <sup>b</sup>

<sup>a</sup> Department of Genetics & Plant Breeding, S.V. Agricultural College, Tirupati-517 502, Andhra Pradesh, India.

<sup>b</sup> Department of Genetics & Plant Breeding, SMGR Agricultural College, Udayagiri-524 226, Andhra Pradesh, India.

## Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

## Article Information

DOI: <https://doi.org/10.9734/ijpss/2025/v37i105788>

## Open Peer Review History:

This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: <https://pr.sdiarticle5.com/review-history/145490>

**Original Research Article**

**Received: 15/08/2025**  
**Published: 17/10/2025**

## ABSTRACT

Castor (*Ricinus communis* L.) is a vital crop had wide range of industrial applications. It is widely used in the production of biodiesel, soaps, inks, varnishes, linoleum, and plasticizers. Understanding genetic diversity is essential for effective breeding strategies. In this study, 40 castor genotypes were evaluated using correlation, principal component analysis (PCA). Correlation analysis revealed that characters with plant height upto primary spike, effective primary spike length, number of capsules on primary spike and 100 seed weight displayed significant and positive correlation of total seed yield. PCA identified five components (with Eigenvalues >1), explaining 85.85% of the total variation. PC-I showed strong positive loadings for number of capsules to

\*Corresponding author: E-mail: kurubachandana05@gmail.com;

primary spike, effective primary spike length, 100 seed weight, total seed weight and number of effective spikes per plant. Considering the mean values of canonical vectors, genotypes ACI-01, ACI-26, and ACI-28 emerged as promising genotypes for use in breeding programmes.

**Keywords:** *Ricinus communis* L.; correlation; principal component analysis; eigenvalues; seed yield.

## 1. INTRODUCTION

Castor (*Ricinus communis* L.,  $2n=2x=20$ ), a premier non-edible commercial oilseed crop, belongs to spurge family, commonly known as Euphorbiaceae. India is the global supplier of castor oil and its by-products. India alone produces 18.79 LMT of castor bean from 10.30 lakh ha cultivated area with an average productivity of 1824 kg ha<sup>-1</sup> (INDIASTAT, 2023-24). In India, castor is chiefly grown in the states of Gujarat followed by Rajasthan, Andhra Pradesh, Telangana, Tamil Nadu, Karnataka and Madhya Pradesh. Castor oil contains substantial quantity of the unusual hydroxylated fatty acid "ricinolate" which enhances its consumption as a lubricant in power engines. It can readily be dissolved in alcohol and transformed into biodiesel even at low temperatures. It is also used to manufacture soaps, printing inks, linoleum, varnishes and plasticizers. Castor oil is a potent laxative and is also a curing agent for skin problems *viz.*, sunburn, crinkles and stretch symbols, *etc.* The by-product of castor oil expeller is the cake which contains vital organic nitrogen (6.4%), P<sub>2</sub>O<sub>5</sub> (2.5%) and K<sub>2</sub>O (1%) including micronutrients for organic farming. The plant stems are used as firewood and to prepare paper pulp by the paper mills. Beyond this, fresh castor leaves are used to rear eri-silkworms, while desiccated leaves are used as an insecticide in agriculture.

The success of plant breeding hinges on the existing genetic variability within the crop (Zheng et al., 2010). Assessing and estimating genetic variability in the germplasm is crucial before initiating any crop improvement program or selecting appropriate breeding techniques. Correlation studies between yield and yield-contributing traits are valuable for breeders to design hybridization programs and evaluate individual plants in segregating populations. To further enhance selection efficiency, multivariate statistical techniques provide a more comprehensive approach. Tools such as Principal Component Analysis (PCA) and cluster analysis are widely used to assess genetic diversity and explore complex trait interactions (Brown-Guedira et al., 2000). Among these, PCA is especially effective, as it reduces large datasets into a few principal components that

explain most of the variation. This allows for the identification of the most influential traits contributing to genetic diversity (Sun et al., 2019). In addition, PCA groups genotypes based on overall performance, aiding in the clear interpretation of genetic diversity (Munir et al., 2020).

## 2. MATERIALS AND METHODS

The present investigation was carried out during *khariif*, 2024 at Agricultural Research Station, Ananthapuramu, Andhra Pradesh. The experimental material utilized for the present study comprised of 40 castor genotypes (37 inbred lines + three varietal checks) sown under rainfed alfisols in alpha lattice design replicated twice. Each genotype was planted in two rows of 6 m length adopting a spacing of 90 cm × 60 cm. All the recommended package of practices were adopted to raise a healthy crop. Observations on five randomly selected plants in each genotype for 12 metric traits *viz.*, days to 50% flowering of primary spike, days to maturity of primary spike, plant height up to primary spike (cm), number of nodes to primary raceme, effective primary spike length (cm), number of capsules per primary spike, number of effective spikes per plant, volume weight (g/100ml), hundred seed weight (g), total seed yield (kg/ha), final plant stand (no./plot) and oil content (%) were studied. Oil content in castor genotypes were estimated by Nuclear Magnetic Resonance Spectrometry (NMR) at Indian Institute of Oilseeds Research (IIOR), Hyderabad. The statistical analysis for various traits was computed using R statistical package.

### 2.1 Statistical Analysis

The statistical procedures *viz.*, correlation (Johnson et al., 1955), principal component analysis (Sneath and Sokal, 1973) were carried out in R Studio (Version 4.4.3).

## 3. RESULTS AND DISCUSSION

### 3.1 Correlation Studies

The basic statistical analysis indicated a considerable degree of variability among the 40 castor r the traits genotypes fo studied (Table 1). Days to 50% flowering of primary spike

registered a significant and positive correlation with days to maturity of the primary spike and negative significant association with volume weight, effective primary spike length (Table 2 and Figs. 1&2). Sadaiah et al. (2021) confirmed significant and positive correlation of days to 50% flowering for effective primary spike. length and number of effective spikes per plant.

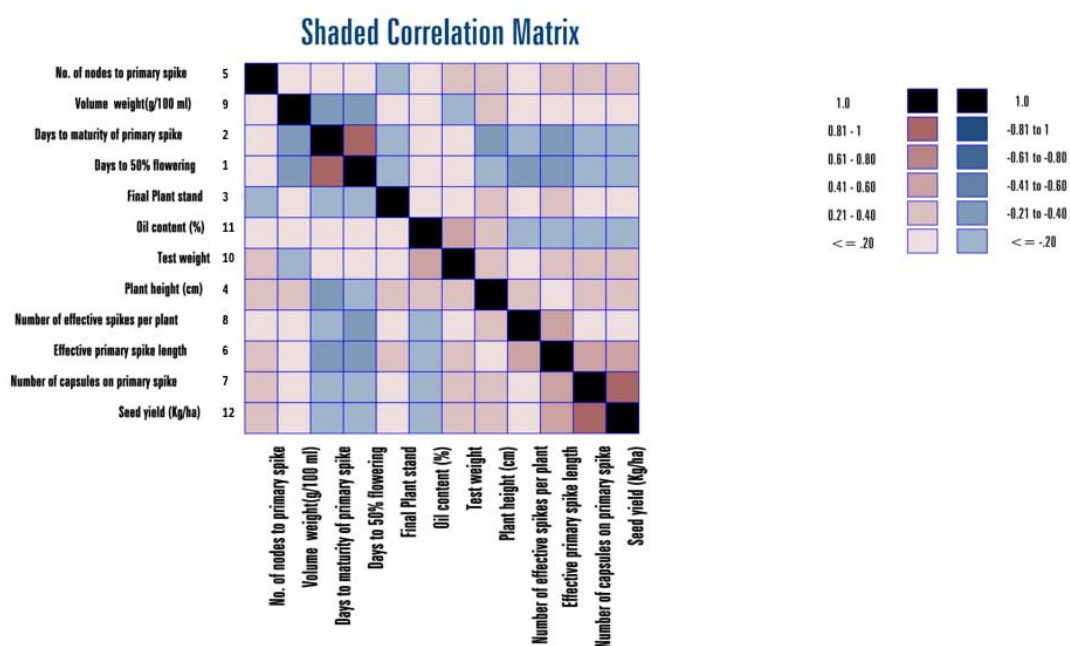
Days to maturity of primary spike exhibited a significantly negative correlation with plant height up to primary spike, effective primary spike length volume weight Reddy et al. (2022) confirmed significant and negative correlation of days to maturity of primary spike with number of effective spikes per plant.

Final plant stand showed significantly and positive association with plant height upto primary spike and effective primary spike length.

Plant height upto the primary spike displayed significantly and positively correlated with the number of nodes up to the primary raceme, number of capsules per primary spike, number of effective spikes per plant, volume weight, 100 seed weight, oil content and total seed yield. Reddy et al. (2022) and Deepak et al. (2024) confirmed significant positive correlation of plant height up to primary spike for number of nodes to primary raceme and number of capsules per spike and oil content.

**Table 1. Trait-wise statistical overview of 40 castor genotypes**

S. No.	Trait	Mean	Min	Max	SEm
1	DFF	55.66	50.00	64.00	1.22
2	DTM	105.80	99.50	114.50	1.37
3	FPS	17.40	15.00	20.00	0.73
4	PLH	47.04	34.30	55.30	3.32
5	NPS	13.27	10.50	15.50	0.57
6	ESL	24.71	15.60	39.90	1.34
7	NCS	51.14	40.40	83.10	2.59
8	NES	3.42	2.10	5.10	0.19
9	VWT	64.53	58.54	71.55	2.29
10	HSW	29.09	22.45	39.45	1.20
11	OCT	47.69	44.00	51.85	1.17
12	TSY	807.32	495.50	1661.40	69.17



**Fig. 1. Phenotypic correlation matrix for yield and its attributes in castor genotypes**

**Table 2. Phenotypic and Genotypic correlations for yield and its attributes in 40 genotypes of castor**

CHARACTER	r	DTM	FPS	PLH	NPS	ESL	NCS	NES	VWT	HSW	OCT	TSY
DFF	r <sub>p</sub>	0.950**	-0.106	-0.198	0.071	-0.251*	-0.102	-0.205	-0.273*	0.108	0.023	-0.040
	r <sub>g</sub>	0.954**	-0.276*	-0.358**	0.179	-0.268*	-0.078	-0.257*	-0.381**	0.152	-0.053	-0.065
DTM	r <sub>p</sub>		-0.086	-0.223*	0.028	-0.249*	-0.183	-0.194	-0.323**	0.073	0.012	-0.094
	r <sub>g</sub>		-0.245*	-0.388**	0.125	-0.279*	-0.166	-0.217	-0.475**	0.102	-0.048	-0.125
FPS	r <sub>p</sub>			0.261*	-0.026	0.246*	0.125	0.149	0.062	0.175	0.185	0.129
	r <sub>g</sub>			0.468**	-0.138	0.407**	0.330**	0.095	0.782**	0.175	0.035	0.147
PLH	r <sub>p</sub>				0.255*	0.197	0.249*	0.301**	0.370**	0.377**	0.339**	0.335**
	r <sub>g</sub>				0.529**	0.236*	0.363**	0.397**	0.415**	0.510**	0.770**	0.419**
NPS	r <sub>p</sub>					0.263*	0.279*	0.194	0.026	0.399**	0.127	0.210
	r <sub>g</sub>					0.341**	0.297**	0.296**	0.080	0.609**	0.161	0.159
ESL	r <sub>p</sub>						0.560**	0.465**	0.174	0.254*	-0.121	0.424**
	r <sub>g</sub>						0.648**	0.504**	0.329**	0.301**	-0.239*	0.482**
NCS	r <sub>p</sub>							0.164	0.113	0.314**	-0.046	0.834**
	r <sub>g</sub>							0.236*	0.091	0.387**	-0.097	0.899**
NES	r <sub>p</sub>								0.194	0.060	-0.150	0.167
	r <sub>g</sub>								0.346**	0.058	-0.318**	0.203
VWT	r <sub>p</sub>									-0.028	0.117	0.051
	r <sub>g</sub>									-0.185	0.195	0.091
HSW	r <sub>p</sub>										0.428**	0.298**
	r <sub>g</sub>										0.511**	0.398**
OCT	r <sub>p</sub>											-0.040
	r <sub>g</sub>											-0.061

\*, \*\* Significant at 5% and 1% levels, respectively

DFF-Days to 50% flowering of primary spike

DTM-Days to maturity of primary spike

FPS - Final plant stand (no. plot<sup>-1</sup>)

PLH-Plant height up to primary spike (cm)

NPS-Number of nodes to primary raceme

ESL- Effective primary spike length (cm)

NCS -Number of capsules per primary spike

NES -Number of effective spikes per plant

VWT- Volume weight (g 100ml<sup>-1</sup>)

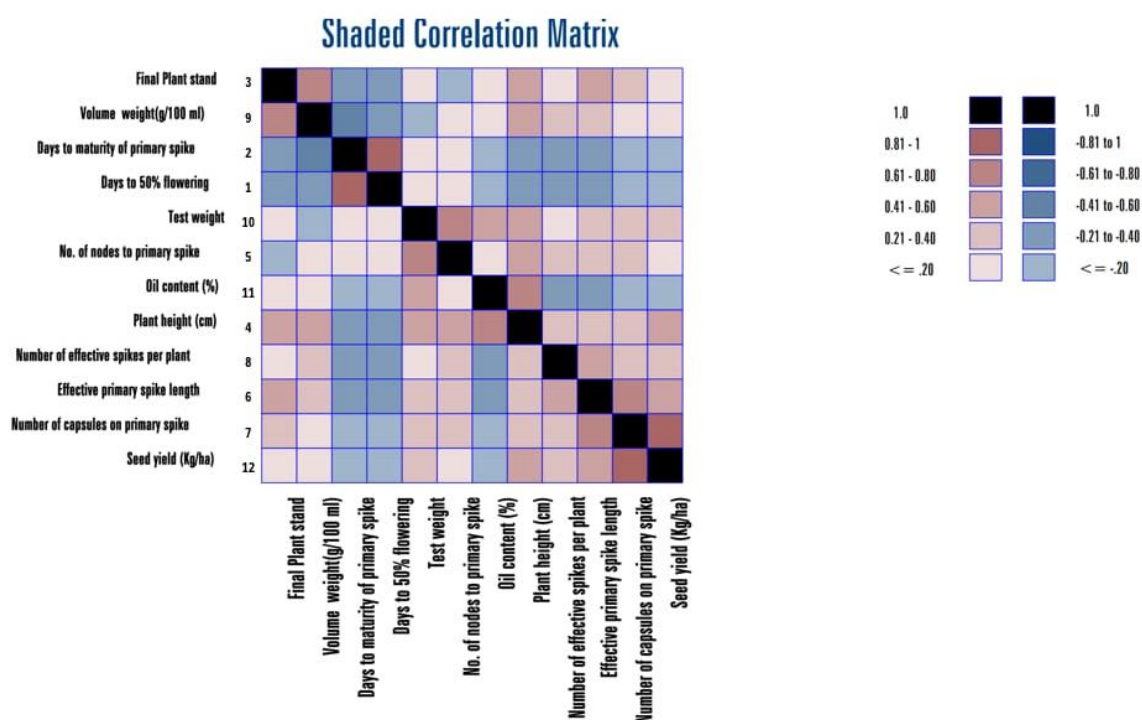
HSW- Hundred seed weight (g)

OCT- Oil content (%)

TSY- Total seed yield (kg ha<sup>-1</sup>)

**Table 3. Principal component analysis of various traits in castor genotypes**

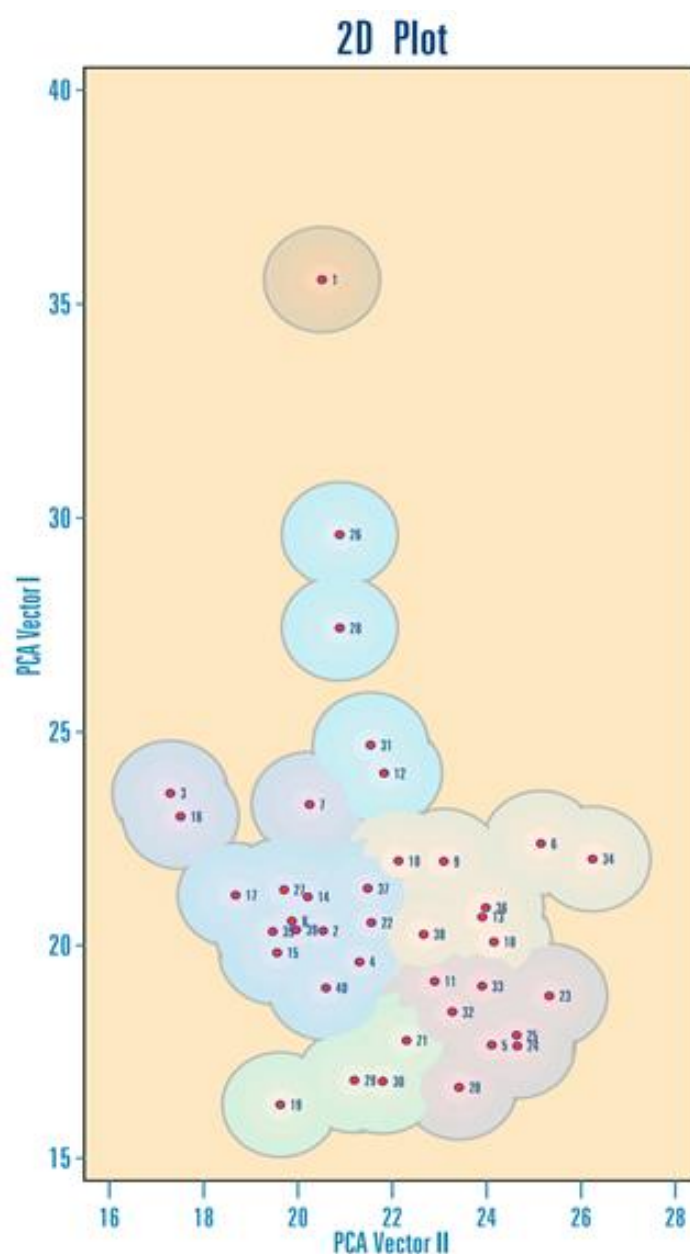
Canonical Roots Analysis (P. C. A.)							
SI. No.		1 Vector	2 Vector	3 Vector	4 Vector	5 Vector	6 Vector
	<b>Eigen Value (Root)</b>	3.75	1.97	1.68	1.24	1.18	0.62
	<b>% Var. Exp.</b>	31.26	16.43	14.00	10.30	9.85	5.20
	<b>Cum. Var. Exp.</b>	31.26	47.70	61.70	72.00	81.85	87.05
1	DFF	0.01	0.13	0.61	0.22	0.26	0.04
2	DTM	-0.08	0.31	0.07	-0.53	0.34	0.66
3	FPS	-0.20	-0.55	-0.04	-0.01	-0.11	0.47
4	PLH	0.18	0.20	-0.36	0.00	-0.59	0.27
5	NPS	0.24	0.30	0.40	0.02	-0.41	0.09
6	ESL	0.38	-0.34	0.08	-0.23	-0.10	0.09
7	NCS	0.44	-0.20	-0.07	0.05	0.21	0.13
8	NES	0.29	0.15	-0.07	-0.61	-0.03	-0.37
9	VWT	-0.16	-0.38	0.36	-0.18	-0.35	-0.01
10	HSW	0.37	0.19	0.13	0.37	-0.11	0.28
11	OCT	-0.41	0.27	-0.25	0.13	-0.10	0.09
12	TSY	0.34	-0.11	-0.32	0.23	0.30	0.06



**Fig. 2. Genotypic correlation matrix for yield and its attributes in castor genotypes**

Number of nodes to the primary raceme registered a significant and positive correlation with effective primary spike length, number of capsules on primary spike and 100 seed weight. Sadaiah et al. (2021) confirmed significant positive association of number of nodes to primary spike for effective primary spike length, number of capsules per primary spike and 100 seed weight.

Effective primary spike length displayed a significant and positive correlation with the number of capsules on primary spike, number of effective spikes per plant, 100 seed weight and total seed yield. Nagarajan et al. (2019) and Movaliya et al. (2018) confirmed significant positive association of effective primary spike length for number of capsules on primary spike, hundred seed weight, number of effective spikes per plant and seed yield.



**Fig. 3. Two -dimensional picture of canonical root analysis**

Number of capsules per primary spike showed a significant and positive correlation with 100 seed weight and total seed yield. Deepak et al. (2024) confirmed significant positive association of number of capsules per primary spike for oil content and seed yield.

Hundred seed weight displayed significant and positive correlation with oil content and total seed yield. Shah et al. (2015) confirmed significant and positive correlation of 100 seed weight for seed yield and oil content.

### 3.2 Principal Component and Cluster Analysis

Principal Component Analysis (PCA) is a multivariate statistical technique commonly used for data comparison, dimensionality reduction and transformation. It simplifies complex datasets by converting correlated variables into a smaller number of uncorrelated principal components, thereby facilitating the identification of patterns and genetic relationships among genotypes. In this study, five out of six principal components were selected with eigen value

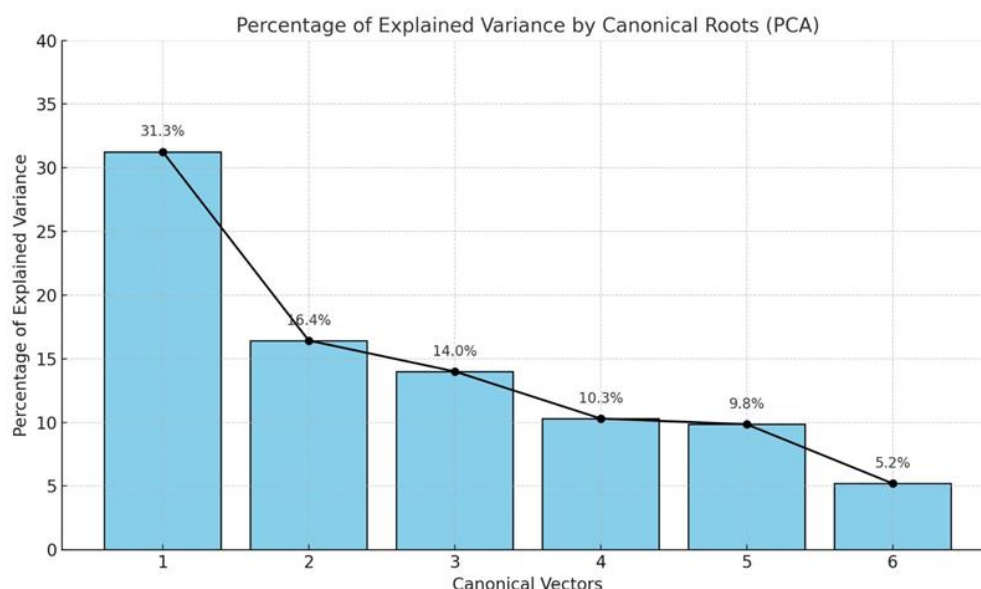
greater than 1 (Table 3 and Fig. 4). Principal Component Analysis revealed that the first five PCs out of six PCs with eigen value >1 contributed 81.85% towards total variability. The PC-I contributed maximum (31.26%), followed by PC-II (16.43%), PC-III (14.00%), PC-IV (10.30%), PC-V (9.85%), PC-VI (5.20%). PC-I was primarily influenced by the traits *viz.*, number of capsules to primary spike, effective primary spike length, 100 seed weight, total seed weight and number of effective spikes per plant all of which exhibited substantial positive loadings, indicating their significant role in explaining the variation captured by PC-I. The traits *viz.*, days to maturity of primary spike, number of nodes to primary spike, oil content exhibited significant high positive loadings in PC-II. Maximum value of positive loadings on PC-III was observed in days to 50% flowering, number of nodes to primary spike and volume weight. In PC-IV, there was maximum positive loadings for days to 50% flowering of primary spike, hundred seed weight and total seed yield. In PC-V, there was maximum positive loadings for days to maturity of primary spike, total seed yield and days to 50% flowering of primary spike. Maximum value of positive loadings on PC-VI was observed in days to maturity of primary spike, final plant stand and hundred seed weight.

The PCA highlighted the extent of variability among the traits, which can be utilized in future breeding programs to enhance yield-related characteristics (Nazir et al., 2013).

The analysis thus identified the maximum contributing variables for diversity are *viz.*, number of capsules per primary spike, effective primary spike length, 100 seed weight, total seed yield. These findings are also confirmation Deepak et al. (2024).

In general, plant breeders retain only those principal components with eigen values greater than one, as these components contribute significantly to the comprehensive variation in the data. The first principal component seizes maximum possible variability, while each subsequent component elucidates progressively less of the remaining variation. In the present study, the first five principal components *viz.*, PC1 (3.75), PC2 (1.97), PC3 (1.68), PC4 (1.24), PC5 (1.18). each had eigen values greater and are more apt that they contribute substantially to the total variation and are considered relevant for further interpretation. Components with eigen values less than one explain less variance than an individual original variable and are therefore considered ancillary in the analysis. Comparable trends have also been observed in the findings of cherukupalli *et.al.* (2021) and Deepak et al. (2024).

The scree plot showing these values are presented in the Fig 4. Hence, the insights gained from Principal Component Analysis can be utilized to derive meaningful and valuable outcomes.



**Fig. 4. Scree plot showing the eigen values of principal components in PCA among 40 genotypes of castor based on PCA scores**

Considering the mean values of canonical vectors, genotypes ACI-01, ACI-26, and ACI-28 emerged as promising genotypes for use in breeding programmes as shown in Fig. 3.

By combining correlation and multivariate analyses, such as PCA, it identifies the most influential traits and promising genotypes, providing a scientific basis for targeted breeding.

The findings can guide breeders in selecting superior genotypes with higher seed yield and oil content, thereby enhancing castor productivity. Overall, this study contributes to the understanding of trait interactions and genetic diversity, supporting sustainable castor cultivation and industrial applications.

#### 4. CONCLUSION

Canonical root analysis accounted for 85.85% by first five canonical vectors towards total genetic divergence of characters based on criteria with eigen values > 1. Considering mean values of canonical vectors, genotypes ACI-01, ACI-26, and ACI-28 emerged as promising genotypes and could be utilized in future breeding programmes.

Correlation studies revealed that characters with plant height upto primary spike, effective primary spike length, number of capsules on primary spike and 100 seed weight displayed significant and positive correlation of total seed yield at both phenotypic and genotypic levels. This indicated possibility of simultaneous selection of all these characters for yield improvement.

#### DISCLAIMER (ARTIFICIAL INTELLIGENCE)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

#### COMPETING INTERESTS

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

#### REFERENCES

Brown-Guedira, G. L., Thompson, J. A., Nelson, R. L., & Warburton, M. L. (2000).

Evaluation of genetic diversity of soybean introductions and North American ancestors using RAPD and SSR markers. *Crop Science*, 40, 815–823.

Cherukupalli, L., & Mukta, N. (2021). An assessment of genetic diversity and combining ability of elite castor genotypes suitable for rainfed conditions. *Electronic Journal of Plant Breeding*, 12(1), 129–136.

Deepak, K. A., Manjunatha, T., Hemalatha, V., & Chary, D. S. (2024). Variability, correlation patterns and principal component analysis (PCA) for seed yield and contributing traits in castor (*Ricinus communis* L.). *Journal of Advances in Biology and Biotechnology*, 27(8), 1217–1227.

INDIASTAT. (2023–2024). Agriculture area and crops growth statistics 2023, 2024 [online]. Available at: <https://www.indiastat.com/data/agriculture> [Accessed 18 June 2024].

Johnson, H. W., Robinson, H. F., & Comstock, R. E. (1955). Estimates of genetic and environmental variability in soybeans.

Movaliya, H. M., Chovatia, V. P., Madariya, R. B., Mungala, R. A., Pipaliya, H. R., & Bhuvu, S. K. (2018). Study of variability and correlation for seed yield and its attributes in castor (*Ricinus communis* L.). *Journal of Pharmacognosy and Phytochemistry*, 7(2), 1474–1477.

Munir, S., Qureshi, M. K., Shahzad, A. N., Nawaz, I., Anjam, M. S., Rasul, S., & Zulfiquar, M. A. (2020). Genetic dissection of interspecific and intraspecific hybrids of cotton for morphoyield and fiber traits using multivariate analysis. *Pakistan Journal of Agricultural Research*, 33, 9–16.

Nagarajan, S., & Viswanathan, P. L. (2019). Genetic divergence analysis of castor (*Ricinus communis* L.). *Electronic Journal of Plant Breeding*, 10(2), 754–760.

Nazir, A., Farooq, J., Mahmood, A., Shahid, M., & Riaz, M. (2013). Estimation of genetic diversity for CLCuV earliness and fiber quality traits using various statistical procedures in different crosses of *Gossypium hirsutum* L. *Bulletin of Agrarian Science*, 43, 2–9.

Reddy, P. S., Priya, P. B., Sadaiah, K., Vanisri, S., & Kumar, C. S. (2022). Genetic variability and character association studies among yield attributing traits in genepool lines of castor (*Ricinus communis* L.). *Biological Forum – An International Journal*, 14(2), 1277–1281.

- Sadaiah, K., Neelima, G., Rani, C. V. D., Rani, V. D., Madhuri, G., Nalini, N., Sujatha, M., Shankar, V. G., Kumar, M. P., Lavanya, C., & Lal, J. J. (2021). Genetic parameters, diversity and character association studies in germplasm lines of castor (*Ricinus communis* L.). *Electronic Journal of Plant Breeding*, 12(4), 1134–1141.
- Shah, S. K., Patel, D. K., Patel, R. M., & Patel, P. S. (2015). Seed and oil quality characteristics of some castor (*Ricinus communis* L.) inbred lines. *International Journal of Agriculture, Environment and Biotechnology*, 8(1), 177–181.
- Sun, Z., Xingfen, W., Zhengwen, L., Qishen, G., Yan, Z., Zhikun, L., Huifeng, K., Jun, Y., Jinhua, W., Liqiang, W., Guiyin, Z., & Zhiying, M. A. (2019). Evaluation of the genetic diversity of fiber quality traits in upland cotton (*Gossypium hirsutum* L.) inferred from phenotypic variations. *Journal of Cotton Research*, 2, 1–8.
- Zheng, L., Qi, J. M., Fang, P. P., Su, J. G., Xu, J. T., & Tao, A. F. (2010). Genetic diversity and phylogenetic relationship of castor as revealed by SRAP analysis. *Wuban Zhiwuxue*, 28, 1–6.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the publisher and/or the editor(s). This publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© Copyright (2025): Author(s). The licensee is the journal publisher. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:

The peer review history for this paper can be accessed here:

<https://pr.sdiarticle5.com/review-history/145490>