



RESEARCH ARTICLE

Principal component analysis for evaluating yield-contributing traits in rapeseed-mustard (*Brassica* spp.) under natural white rust infection

Shivansh Singh¹, Prabhakar Verma¹, Vijay Kumar Kushwaha¹, Gagan Kumar², Hemant Kumar Yadav³ & Arun Kumar^{1*}

¹Department of Genetics and Plant Breeding, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya 224 229, Uttar Pradesh, India

²Department of Plant Pathology, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya 224 229, Uttar Pradesh, India

³Department of Molecular Biology & Biotechnology, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya 224 229, Uttar Pradesh, India

*Correspondence email - arungpb@nduat.org

Received: 25 July 2025; Accepted: 17 June 2026; Available online: Version 1.0: 18 August 2026

Cite this article: Shivansh S, Prabhakar V, Vijay KK, Gagan K, Hemant KY, Arun K. Principal component analysis for evaluating yield-contributing traits in rapeseed-mustard (*Brassica* spp.) under natural white rust infection. Plant Science Today. 2026; 13(sp5): 1-5. <https://doi.org/10.14719/pst.10822>

Abstract

The present investigation was conducted during the 2023 rabi season at Acharya Narendra Deva University of Agriculture and Technology, Ayodhya, to assess the genetic variability and identify key yield-contributing traits in 100 rapeseed mustard (*Brassica* spp.) genotypes under natural white rust infection using principal component analysis (PCA). Eleven quantitative traits were evaluated and PCA revealed that five principal components (PCs) had eigenvalues greater than 1, cumulatively explaining 69.51 % of the total genetic variation. The first principal component (PC1) accounted for the highest variance (17.2 %) and was primarily influenced by harvest index, seed yield per plant and days to maturity. The second principal component (PC2) contributed 16.5 % of the variance and was defined by disease intensity, days to 50 % flowering, test weight and number of siliquae per plant. Subsequent components also showed significant contributions from key agronomic traits such as biological yield, silique length and seeds per silique. The analysis helped distinguish genotypes such as PAB 9511, GIRIRAJ, JAWAHAR MUSTARD-3 and EC 399301 as superior based on their high contributions to yield traits. These findings provide valuable insights for selection and breeding strategies aimed at improving yield potential and disease resilience in rapeseed mustard.

Keywords: genetic diversity; PCA; rapeseed-mustard; rust

Introduction

Rapeseed and mustard are among the most important oilseed crops within the *Brassica* group, serving as a key source of edible oil for a large segment of the global population (1). These crops are widely cultivated across both developing and developed countries and they are known for their rich genetic diversity. The wide variation found in *Brassica* germplasm is a valuable resource for breeders and researchers alike (2, 3). Gaining a clear understanding of the genetic differences among various genotypes is a crucial starting point for any breeding program aiming to develop improved, high-yielding varieties (4, 5). Wild species and traditional landraces, which have evolved under diverse environmental conditions, offer a treasure trove of genes that can help strengthen future crop improvement efforts (6–8). However, despite the abundance of genetic resources available, only a small fraction has been tapped into, leading to a concerning reduction in genetic diversity among today's commercial cultivars (9).

Breeders working with rapeseed-mustard (*Brassica* × *juncea* (L.) Czern.) often assess numerous traits, but not all of them are equally useful for differentiating among genotypes.

In such scenarios, principal component analysis (PCA) becomes a powerful tool, helping to identify meaningful patterns while removing redundant or overlapping information (10). As a multivariate statistical method, PCA simplifies complex datasets by condensing them into a smaller number of principal components (PCs), each defined by eigenvectors and corresponding scores (11). This approach is especially effective when dealing with several interrelated quantitative traits, making it easier to understand the underlying variation within the germplasm (12, 13).

Principal component analysis works by converting the original data into a set of new, uncorrelated variables called PCs that capture most of the variation present in the dataset (14, 15). Using just a few of these components, each genotype can be effectively summarised by scores, rather than dealing with raw values for every trait. The true strength of PCA lies in its ability to pinpoint which combinations of traits explain the majority of the variability, making it much easier to interpret complex data and visualise genetic relationships. Traits that have higher loading values whether positive or negative play a key role in distinguishing among different germplasm lines.

Recognising the usefulness of PCA, a study was undertaken to apply PCA to yield and yield-related traits in a collection of *B. juncea* (rapeseed-mustard) accessions. This analysis helped uncover the key traits contributing most to variation across the genotypes and provided valuable insights for identifying superior lines. The findings are expected to support breeding efforts aimed at improving seed yield and oil content in future cultivars.

Materials and Methods

This experiment was carried out at Students Instructional Farm (26.542031° N latitude and 81.82258° E longitude), College of Agriculture, Acharya Narendra Deva University of Agriculture and Technology (26°28' N latitude and 82°57' E longitude, 113 m above mean sea level), Ayodhya, Uttar Pradesh, during November, rabi 2023. The climate of the region is tropical and humid, with a mean annual rainfall of approximately 1120–1140 mm. The experimental

material consists of 100 rapeseed-mustard genotypes (Table 1). Various Indian and exotic collections, along with 2 checks, namely, Varuna and EC-399294, were used in the present study. Crop was sown in the 2nd week of November 2023 and was ready for harvest in the mid of March 2024. Genotypes were sown in Alpha lattice design by Patterson and Williams, 1976. The experimental field was divided into 2 replications with 10 blocks in each replication and 1 block consisted of 10 genotypes. Each block consists of 10 rows of 2 m in length, each with a spacing of 50 cm between rows and 1 m between the blocks and 2 m between the replications and a planting space of 45 x 30 cm. Five random plants were tagged from each block for data collection. A total of 11 quantitative traits viz., days to 50 % flowering, days to maturity, plant height (cm), number of siliquae per plant, seeds per silique, silique length, test weight, harvest index, biological yield per plant, seed yield per plant and disease intensity. The observations recorded were statistically analysed using RStudio 2025.05.1 + 513 software.

Table 1. List of 100 rapeseed-mustard genotypes used in the study

Sl. No.	Name	Sl. No.	Name
1	RADHIKA-1	51	PAB 9511
2	NDR 2005-1	52	GIRIRAJ
3	LAXMI	53	PM-25
4	DRMR-150-35	54	NIGRA
5	UNKNOWN-3	55	JAWAHAR MUSTARD-3
6	GUJRAT MUSTARD-2	56	CS-60
7	EC 399294	57	KRISHNA
8	RH-749	58	EC 399301
9	SHIVANI	59	LACMA-SP523N5
10	RB-50	60	PM-29
11	RCH-1	61	VARUNA-3
12	RGN-236	62	PL-6
13	RGN-13	63	RH-819
14	PM-25	64	RSMY 43
15	CS-58	65	RH-749
16	KANTI	66	PBR-97
17	RH-761	67	JATHI
18	JAGANNATH	68	RADHIKA
19	PUSA AGRANI	69	RGN-229
20	ARAWALI	70	TARAMIRA
21	GEETA	71	VARUNA
22	SBG-16-2	72	PM-24
23	VAIBHAV	73	PM-27
24	PM-30	74	RH-406
25	NRCHB-101	75	MAYA
26	SBG-17-32	76	PUSA TARAK
27	AIRCYS-05-02	77	PM-21
28	HEERA	78	PBR-91
29	JAWAHAR MUSTARD-1	79	GUJRAT MUSTARD-3
30	BHAGIRATHI	80	BRIJRAJ
31	UNKNOWN-1	81	RH-725
32	CS-60	82	RH-781
33	NRCDR-2	83	PM 31
34	RGN-73	84	JAWAHAR MUSTARD-2
35	DONSKAZA	85	VARDAN
36	RGN-48	86	CS-52
37	BR-40	87	PM-22
38	KRANTI	88	GSH 1707
39	C-724	89	PM-28
40	PUSA VIJAY	90	CS-56
41	CS-54	91	PUSA BAHAR
42	PT-3	92	PUSABOLD
43	RH-30	93	PM-26
44	RH-125	94	VARUNA
45	RADHIKA-3	95	PBR-210
46	RH-119	96	UNKNOWN-2
47	GIRIRAJ	97	PM-30
48	NDR-8501	98	EC 399302
49	RLC-3	99	GIRIRAJ-1
50	PDL-1	100	MASC-1

Results and Discussions

The results of the PCA revealed the extent of genetic variation among the genotypes based on various agro-morphological traits. Components with eigenvalues greater than 1 and accounting for more than 4 % of the total variation were considered significant, following the criterion suggested from a previous study (16). The PCA effectively captured the genetic diversity present among the rapeseed–mustard genotypes under study. Eigenvalues represent how much variation each PCs explains in the dataset. Meanwhile, the coefficients of the eigenvectors indicate how much each original trait contributes to its respective PCs. However, studies have shown that there are no standardised statistical tests to determine the significance of eigenvalues or their associated coefficients. (Fig. 1, Table 2) (17). Overall, these values help highlight which traits play a major role in explaining the observed variation among genotypes.

The characters coming together in different PCs explaining the variability show the tendency to remain together and must be taken into consideration during the exploitation of these characters in the breeding program (18). Principal component analysis revealed that out of 11 characteristics studied, only five PCs exhibited more than 1 eigenvalue and showed about 69.5 % cumulative variability among the traits studied while the other components

were rejected because they have eigenvalues less than 1. So, these five PCs were given due importance for further explanation. Out of the five PCs exhibiting more than 1.00 eigenvalue, PC1 had the highest variability (17.2%), followed by PC2 (16.5%).

The first principal component PC1 was mainly contributed by three characters viz., harvest index (24.2 %), seed yield per plant (18.8 %), days to maturity (16.3 %). The second principal component (PC2) was contributed mostly by four characters like disease intensity (22.5 %), days to 50 % flowering (16.8 %), test weight (16.8 %), number of siliquae per plant (16.4 %). Seed yield per plant and yield attributing traits i.e., number of siliquae per plant, test weight and siliqua length, positively affect the PC2. This indicates the increase in seed yield per plant and yield attributing traits i.e., number of siliquae per plant, test weight and siliqua length as a result of late flowering. To be sure, improving a given yield trait will direct the improvement of other yield traits collected in the same PC as long as they have the same positive effect.

The third principal component (PC3) is positively associated with harvest index (25.4 %), seeds per siliqua (22.5 %), seed yield per plant (17.9 %) and days to 50 % flowering (10.7 %). The fourth principal component (PC4), is positively associated with biological yield per plant (28.8 %), siliqua length (22.4 %) and disease intensity

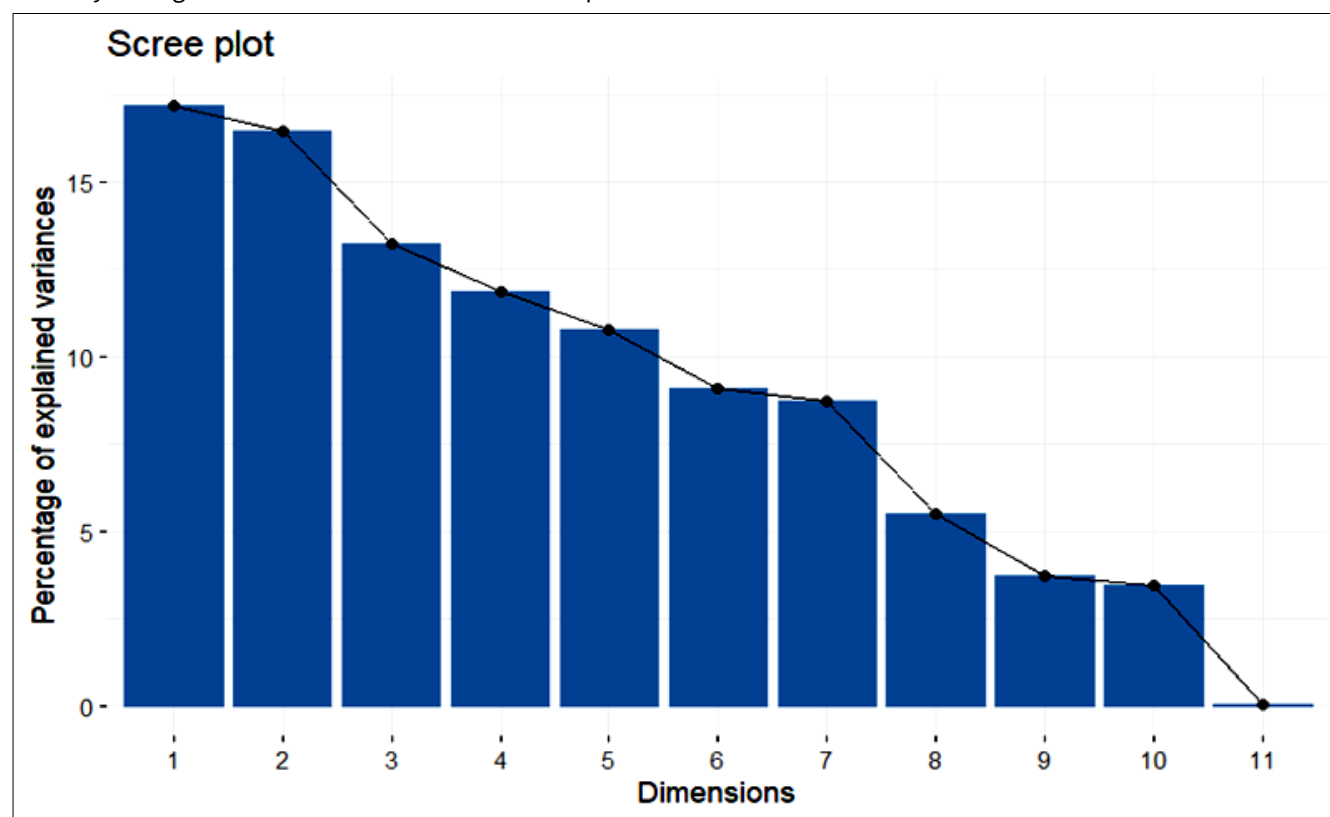


Fig. 1. Scree plot diagram of eigenvalues constructed on 11 traits recorded in 100 rapeseed genotypes.

Table 2. Eigenvalue, proportion and cumulative variation of analysed components

Principal components	Eigenvalue	Proportion variance (%)	Cumulative variance (%)
PC1	1.891	17.193	17.193
PC2	1.811	16.467	33.66
PC3	1.455	13.225	46.885
PC4	1.304	11.859	58.744
PC5	1.184	10.763	69.507
PC6	0.999	9.084	78.591
PC7	0.96	8.729	87.32
PC8	0.603	5.481	92.801
PC9	0.409	3.718	96.519
PC10	0.38	3.458	99.977
PC11	0.003	0.023	100

(13.4%). The fifth principal component (PC5), is positively associated with biological yield per plant (40.4 %), siliqua length (25.5 %), seed yield per plant (12.2 %) (Fig. 2, Table 3). Through PCA we could identify the characters responsible for genotypic variation within the group. Five PCs with eigenvalues greater than 1 explained 69.5 % of the total variance (19). It indicates that the identified traits within the axes exhibited great influence on the phenotype of germplasm lines. Principal component analysis has been used by various researchers for the characterisation of different rapeseed-mustard germplasm lines (20–22). Principal component analysis helps us to identify the characters that have a great impact on the phenotype of different genotypes of rapeseed-mustard and this is very much important in the selection procedure of the breeding programme.

Conclusion

The first principal component (PC1) was positively contributed mainly by two characters viz., harvest index, seed yield per plant and days to maturity. The second principal component (PC2) was contributed by four characters like disease intensity, days to 50 % flowering, test weight and number of siliquae per plant. On the basis of PC analysis as depicted in Fig. 2, the genotypes PAB 9511, GIRIRAJ,

JAWAHAR MUSTARD-3, EC 399301, are showing maximum contribution for yield attributing traits.

Acknowledgements

We would like to express our gratitude to all of our colleagues and who helped us complete this study.

Authors' contributions

SS contributed to the preparation of the original draft manuscript. VKK was involved in investigation, formal analysis and data correction. PV contributed to data correction. GK was responsible for formal investigation and provided suggestions. HKY contributed to the review and correction of the manuscript. AK was involved in conceptualisation, monitoring, review and editing of the original draft. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interest to declare.

Ethical issues: None

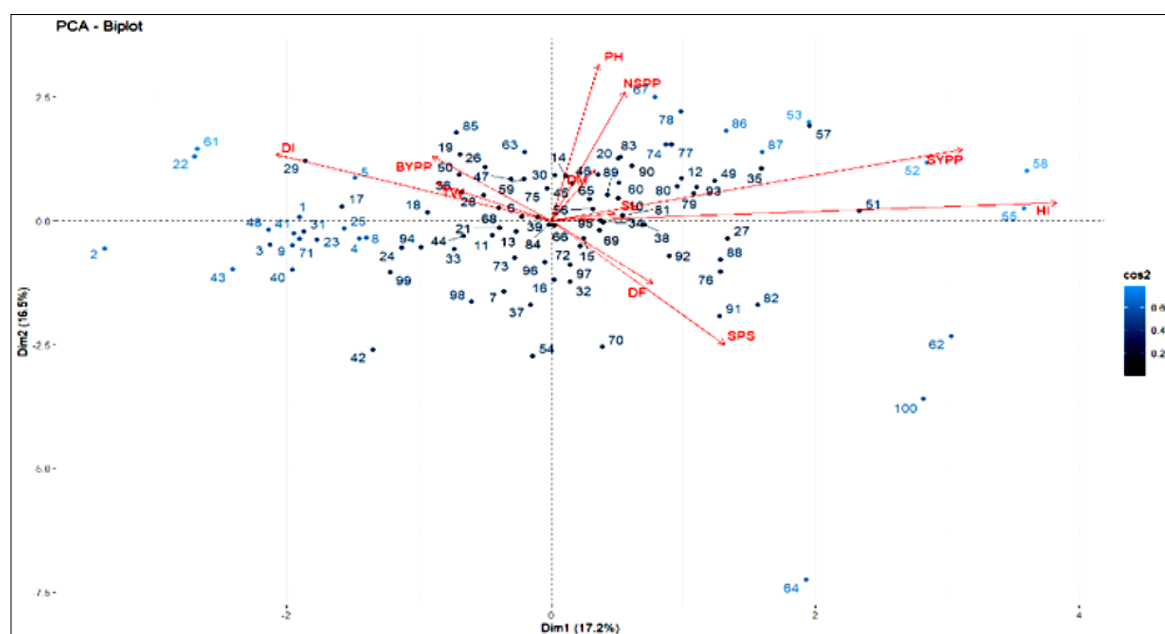


Fig. 2. Biplot distribution of 100 rapeseed-mustard genotypes and studied traits on principal component axes PC1 and PC2.

Table 3. Principal components for studied traits in 100 rapeseed-mustard genotypes

Characters	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11
DF	5.698	16.858	10.742	3.921	2.279	8.559	15.746	0.056	2.194	33.768	0.18
DM	16.331	3.201	5.077	6.384	7.255	14.465	1.036	27.205	5.031	13.858	0.155
PH	12.411	11.368	9.54	2.44	0.318	11.789	0.231	17.369	32.815	1.714	0.006
NSPP	7.469	16.442	1.143	3.601	9.372	17.887	0.695	36.89	6.074	0.419	0.008
SL	1.856	6.656	0.723	22.433	25.523	12.203	0.318	0.486	18.63	10.867	0.306
SPS	5.393	3.167	22.503	10.116	0.358	11.689	3.463	12.776	26.026	4.477	0.031
TW	5.53	16.847	1.963	0.003	0.59	1.84	67.536	1.12	4.324	0.238	0.009
HI	24.224	0.216	25.418	1.093	0.422	0.939	0.439	1.761	1.746	0.015	43.726
BYPP	2.016	1.378	2.358	28.86	40.417	0.869	0.273	1.162	3.053	0.616	18.997
SYPP	18.786	1.339	17.915	7.802	12.216	0.219	0.01	0.98	0.027	4.31	36.395
DI	0.285	22.528	2.619	13.346	1.251	19.542	10.252	0.195	0.079	29.717	0.187

DF : days to 50 % flowering; DM : days to maturity; PH : plant height; NSPP : number of siliquas per plant; SL : siliqua length; SPS : seeds per siliqua; TW : test weight; HI : harvest index; BYPP : biological yield per plant; SYPP : seed yield per plant and DI : disease intensity.

References

- Shekhawat K, Rathore SS, Premi OP, Kandpal BK, Chauhan JS. Advances in agronomic management of Indian mustard (*Brassica juncea* L.): An overview. *Int J Agron*. 2012;2012:408284. <https://doi.org/10.1155/2012/408284>
- Chauhan JS, Tyagi MK, Singh D, Kumar PR. Status of rapeseed-mustard research in India. *J Oilseed Brassica*. 2020;11(2):77–89.
- Thakur AK, Kumar S, Sharma N, Kumari S. Genetic diversity in Indian mustard (*Brassica juncea* L.) using morphological and molecular markers. *Int J Curr Microbiol Appl Sci*. 2019;8(2):1952–62.
- Singh KH, Singh VV, Meena PD, Singh D. Genetic variability and correlation studies in Indian mustard (*Brassica juncea* L.) under late sown conditions. *J Oilseed Brassica*. 2016;7(1):79–84.
- Yadava DK, Yadava P, Vasudev S, Singh N. Genetic improvement of rapeseed-mustard for yield and quality traits through conventional and molecular approaches. *Indian J Genet Plant Breed*. 2022;82(3):273–85.
- Das S, Chattopadhyay C, Chattopadhyay A. Exploring genetic diversity in Indian mustard (*Brassica juncea*) using morphological and SSR markers. *J Oilseed Brassica*. 2014;5(1):18–23.
- Warwick SI, Francis A, Al-Shehbaz IA. Brassicaceae: Species checklist and database on CD-ROM. *Plant Syst Evol*. 2006;259(2–4):249–58. <https://doi.org/10.1007/s00606-006-0422-0>
- Atri C, Akhtari J, Bishnoi SK, Banga SS. Broadening the genetic base of *Brassica juncea* by introgression from *B. carinata* and *B. nigra*. *Genet Resour Crop Evol*. 2020;67(6):1315–25.
- Pradhan AK, Sodhi YS, Mukhopadhyay A, Arumugam N, Pental D. Breeding high yielding hybrid mustard in India: Past, present and future. *Indian J Genet Plant Breed*. 2012;72(2):141–50.
- Neeru, Kumari R, Meena ML, Sharma S, Saini M. Genetic diversity analysis using principal component analysis in Indian mustard (*Brassica juncea* L. Czern. & Coss.). *J Oilseed Brassica*. 2024;15(1):54–60.
- Patel AD, Patel NB, Chauhan RM. Principal component analysis for yield and yield components under timely sown, drought and late-sown conditions in Indian mustard (*Brassica juncea* L.). *J Oilseed Brassica*. 2024;15(1):61–6.
- Choudhary R, Gupta K, Yadav R, Bhardwaj R, Chaturvedi SK. Unravelling the genetic variability in *Brassica juncea* germplasm for morphological and biochemical traits to identify suitable donors. *J Oilseeds Res*. 2023;40(Special Issue):59–65. <https://doi.org/10.56739/g1vzvq62>
- Shyam C, Tripathi MK, Tiwari S, Tripathi N. Genetic components and diversity analysis in Indian mustard (*Brassica juncea* (Linn.) Czern. & Coss.) based on different morpho-physiological traits. *Curr J Appl Sci Technol*. 2021;40(20):34–57. <https://doi.org/10.9734/cjast/2021/v40i2031462>
- Yadav S, Verma A, Singh R. Study on genetic variability and principal component analysis in Indian mustard (*Brassica juncea* L. Czern. & Coss.). *Int J Biochem Plant Breed*. 2025;9(5 Suppl):4284–90. <https://doi.org/10.33545/26174693.2025.v9.i5Sa.4284>
- Patterson HD, Williams ER. A new class of resolvable incomplete block designs. *Biometrika*. 1976;63:83–90. <https://doi.org/10.1093/biomet/63.1.83>
- Singh T, Gupta SK, Sharma A, Sheera A, Bhatt B. Genetic characterization studies of F₂ population of Indian mustard × East European mustard through principal component analysis. *J Oilseed Brassica*. 2024;15(1):143–5.
- Dinno NC, Villéger S, Peres-Neto PR. PCATest: Permutation-based significance testing for principal component analysis. *Methods Ecol Evol*. 2022;13(5):1025–35.
- Khan Hasan MM, Kundu BC, Ghosh S, Hasan GN, Anik MRH, Islam M, et al. Genetic variability, traits interactions and principal component analysis of broad-leaf mustard group (*Brassica juncea* L.) to focus on the yield improvement. *Int J Agric Sci Food Technol*. 2023;9(4):104–10. <https://doi.org/10.17352/2455-815X.000200>
- Sharma H, Kumar A, Singh VV, Meena HS, Sharma P, Rai PK. Variability and genetic diversity study based on agro-morphological traits in a diverse set of Indian mustard (*Brassica juncea* L.). *J Oilseed Brassica*. 2021;12(2):75–85.
- Neeru, Thakral RA, Singh A. Evaluation and classification of Indian mustard (*Brassica juncea* L.) genotypes using principal component analysis. *J Oilseed Brassica*. 2024;15(1):167–74.
- Choudhary K, Kharb RPS, Meena DK. Genetic diversity and principal component analysis in rapeseed-mustard germplasm. *J Oilseeds Res*. 2023;40(3):145–51. <https://doi.org/10.56739/p5j5z624>
- Ghosh S, Avinash H, Dubey N, Sharadhi GP, Danalakoti K, Sachan S, et al. PCA-based multivariate segmentation of variance in Indian mustard (*Brassica juncea* L.). *SABRAO J Breed Genet*. 2023;55(6):1963–71. <https://doi.org/10.54910/sabao2023.55.6.10>

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonpublishing.com/journals/index.php/PST/open_access_policy

Publisher's Note: Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Indexing: Plant Science Today, published by Horizon e-Publishing Group, is covered by Scopus, Web of Science, BIOSIS Previews, Clarivate Analytics, NAAS, UGC Care, etc
See https://horizonpublishing.com/journals/index.php/PST/indexing_abstracting

Copyright: © The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited (<https://creativecommons.org/licenses/by/4.0/>)

Publisher information: Plant Science Today is published by HORIZON e-Publishing Group with support from Empirion Publishers Private Limited, Thiruvananthapuram, India.