



Plant Archives

Journal homepage: <http://www.plantarchives.org>

DOI Url : <https://doi.org/10.51470/PLANTARCHIVES.2026.v26.no.2.076>

MULTIVARIATE ANALYSIS OF GENETIC AND MORPHOLOGICAL DIVERSITY IN *Brassica napus* L. GERMPLASM

Parul Pandey, Shweta*, Majiga Revanth Kumar, Sandeep Varma, Nitin Kumar, Nikhil Kumar, Chandragupt Maurya, Amit Kumar, and Shivam Tripathi

Department of Genetics and Plant Breeding, CSA University of Agriculture and Technology, Nawabganj, Kanpur, India

*Corresponding author E-mail: shweta@csauk.ac.in

(Date of Receiving-31-05-2026; Date of Revision-15-08-2026; Date of Acceptance-27-08-2026)

ABSTRACT

To improve the new cultivars of rapeseed (*Brassica napus* L.), effective tools for tracking trait connections in a breeding program are needed. The study was conducted in Rabi 2024–2025; the 36 rapeseed genotypes were assessed using a randomized complete block design with three replications. Significant variations were found across all traits using analysis of variance. The estimates of genetic diversity of environmental coefficient of variation (ECV), genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability, genetic advance (GA), and genetic advance as percent of mean (GAM) for the studied traits. The genotypes were grouped into eight clusters. Cluster I contained the maximum number of genotypes (14), and clusters V, VI, VII and VIII contained the lowest (1). Cluster 3 had the highest intra-cluster distance (8.34), followed by Cluster 2 (7.79), Cluster 1 (7.34), and Cluster 4 (4.48). Clusters 5, 6, 7, and 8 had the shortest intra-cluster distance (0.00). According to principal component analysis (PCA), the first four principal components could account for 71.4% of the overall variation among the attributes. For rapeseed breeding, the data on yield and agronomic trait changes found in this study provided helpful parents. Additionally, the results of this study are highly beneficial for advancing any breeding initiatives for oilseed rape to increase yield.

Key words: Analysis of variance, *Brassica napus*, Genetic diversity, Principal component analysis

Introduction

In India and around the world, *Brassica napus* L., commonly known as canola or rapeseed, is an essential oilseed crop due to its high yields of seed, oil content, nutritional value, and industrial uses. Along with rapeseed, the genus *Brassica* is a member of the Brassicaceae or Cruciferae family. Originating from *Brassica rapa* (syn. *campestris*) (AA, $2n = 20$) and *Brassica oleracea* L. (CC, $2n = 18$), it is an amphidiploid species (AACC, $2n=38$) (Nagaharu, 1935; Labana *et al.*, 1993). India's edible oil economy depends heavily on this crop, which has a suitable fatty acid composition and an oil content of 35% to 45% (Shehzad *et al.*, 2015). Oilseed rape (*B. napus* ssp. *napus*) is currently the second most important oilseed crop worldwide after soybeans and the most significant source of vegetable oil in Europe. In addition

to serving as the foundation for energy, oil and fat also contain fat-soluble vitamins A, D, E, and K. The oil cake is used as a very good animal feed and fertiliser for many crops because it contains proteins with high biological value and useful amounts of calcium and phosphorus. In terms of acreage and production, rapeseed, along with mustard, is today the third most important edible oil crop in the world, behind cotton and soybeans. Canada, China, northern Europe, and the Indian subcontinent are major producers (Parvin *et al.*, 2019).

However, the current breeding material has a rather shallow genetic background because of its restricted geographic range and vigorous breeding. A concentration on oil and seed quality characteristics has further degraded the gene pool of elite oilseed rape breeding material. Consequently, this essential crop's genetic variability is

limited in several traits that are crucial for breeding. Therefore, it is essential to create new rapeseed types with a high seed yield. Seed yield is a multifaceted characteristic that ultimately produces a highly variable yield structure (DIEPENBROCK, 2000). It is the outcome of morphological and physiological processes that take place over the course of phenological development. Indirect selection for traits that affect yield can be a successful strategy in a rapeseed breeding program because yield components play a major role in the plant's final yield performance (SADAT *et al.*, 2010; MARJANOVIC-JEROMELA *et al.*, 2011). The kind of model utilised to analyze the data is one of the key elements in comprehending yield constraints.

Characterisation of germplasm through morphological traits is a fundamental step in understanding the diversity available within a crop species. It enables the identification of promising genotypes and supports the efficient use of genetic resources in breeding programmes. Evaluation of morphological, agronomic, physiological, and genetic traits helps in selecting superior accessions for the development of improved cultivars. Germplasm originating from different geographical regions often possesses a broad range of variability, providing breeders with valuable material for enhancing productivity, adaptability, and other desirable characteristics. The effectiveness of a breeding programme largely depends on the amount of genetic variation present in the breeding population. Therefore, assessing the extent and nature of variability among genotypes is essential for successful selection. Statistical parameters such as the genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability, and genetic advance are commonly employed to estimate the magnitude of genetic diversity and the potential response to selection. These parameters help distinguish genetic effects from environmental influences and provide insights into the inheritance patterns of important traits, thereby aiding the formulation of efficient crop improvement strategies.

Assessment of genetic diversity among genotypes is essential for the effective utilization of germplasm in crop improvement programmes. Mahalanobis' D^2 analysis is a widely used multivariate statistical approach for quantifying genetic divergence among genotypes and was further elaborated by Rao (1952). This method has been extensively employed in various crops to identify genetically diverse parents for hybridization programmes.

Conventional breeding approaches, including single-plant selection from locally adapted cultivars and pedigree selection from inter-varietal crosses, have played a significant role in improving seed yield, seed size, oil

content, and earliness. The success of such breeding programmes depends on the availability of genetic variability and a thorough understanding of the genetic relationships among breeding materials. Therefore, the collection, characterization, and utilization of plant genetic resources are fundamental components of crop improvement efforts. Multivariate techniques such as principal component analysis (PCA) and cluster analysis are valuable tools for evaluating genetic diversity and identifying suitable parental combinations. These methods facilitate the classification of genotypes based on multiple traits and help breeders select genetically diverse parents for developing superior progenies. *Therefore*, the present study was undertaken to assess the extent of genetic variation among 36 *Brassica napus* genotypes using Mahalanobis' D^2 statistics and principal component analysis.

Multivariate analysis approaches are used to determine genetic divergence when working with a wide range of germplasm lines, regardless of the type of data, such as morphological, biochemical, or molecular markers. Principal component analysis (PCA) and cluster analysis are the most often utilized algorithms for this goal by different researchers (Ciancaleoni *et al.*, 2014). This research was conducted to highlight the significance of genetic diversity in enhancing *Brassica napus* for increased yields and oil content. The germplasm was assessed to gauge the extent of variation across different traits and to identify diverse and promising lines that could potentially be used to develop a population with improved resistance and greater yield potential.

Materials and Methods

Plant materials

A total of thirty-six genotypes of *Brassica napus*, including one check variety (BM-91), were obtained from the National Bureau of Plant Genetic Resources (NBPGR), New Delhi (Table 1), and evaluated in the present study. In addition, local landraces were collected from the Banda region to assess their adaptability to regional agro-climatic conditions and to identify potential farmer-preferred traits.

Experimental Design and Observations Recorded

The experiment was conducted during the Rabi season of 2024–25 at the Student Instructional Farm (SIF), Chandra Shekhar Azad University of Agriculture and Technology, Kanpur, Uttar Pradesh, India. The trial was laid out in a Randomized Complete Block Design (RCBD) with three replications. Each genotype was sown in a single row of 2 m length, maintaining a spacing of 60 cm between rows and 15 cm between plants. Five

Table 1: Details of the 36 *Brassica napus* genotypes evaluated during Rabi 2024–25.

S. No.	Genotype/Accession
1	EC-338961
2	IC-399682
3	IC-491586
4	EC-657053
5	EC-338979
6	EC-358987
7	EC-657049
8	EC-400804
9	EC-414279
10	EC-358975
11	EC-386738
12	EC-338978
13	EC-414271
14	EC-414271
15	EC-339016
16	IC-399681
17	BM-91
18	EC-389916
19	EC-657036
20	IC-339862
21	EC-338995
22	EC-338968
23	IC-399811
24	EC-339000
25	IC-278080
26	EC-338977
27	EC-339015
28	EC-414273
29	EC-338981
30	EC-414276
31	IC-279074
32	EC-338993
33	EC-339001
34	EC-338985
35	EC-338994
36	EC-338999

plants were randomly selected and tagged in each plot for recording observations. Data were recorded on eighteen quantitative traits, namely days to 50% flowering, days to maturity, plant height (cm), number of primary branches per plant, number of secondary branches per plant, length of main raceme (cm), number of siliquae per main raceme, number of siliquae per plant, number of seeds per siliqua, 1000-seed weight (g), oil content (%), seed yield per plant (g), and other relevant agronomic traits.

Statistical analysis

To assess the extent of variability among the studied genotypes, principal component analysis (PCA) was

Table 2: Analysis of variance of Randomized Complete Block Design (RCBD) for 12 Traits in *Brassica napus* genotypes.

Source of variation	R	T	E
Degree of Freedom	1.84	33.16	1.78
Days to 50% flowering	4.78	5.53**	3
Days to maturity	10.17	869.50**	10.13
Plant height (cm)	67.51	258.00**	7.34
Length of main raceme (cm)	0.25	0.41**	0.46
No. of primary branches	0.49	0.55**	0.37
No. of secondary branches	153.01	236.92**	6.63
No. of siliquae on main raceme	191.82	1047.31**	49.96
No. of siliquae per plant	11.32	13.57**	6.25
Average seed per siliquae	0.27	7.75**	1.31
Seed yield per plant	0	1.36**	0.31
1000 seed weight	0.76	5.70**	4.71
Oil content (%)	1.84	33.16**	1.78

R: Replications; T: Treatments; E: Error; **significant at 1% level

performed using the mean values of the recorded traits. Genotypes were further grouped using K-means and Tocher's clustering methods. Principal component analysis was carried out based on the correlation matrix of the original trait data following the procedure described by Jolliffe (1986). The statistical analysis was done using R software (version 4.5.0). The 'variability', 'ggplot2', and 'factoextra' packages were used for PCA analysis. Hierarchical clustering and dendrograms were generated with the 'hclust' function to identify genetic diversity in the studied materials.

Results

Analysis of Variance

Analysis of variance (ANOVA) revealed highly significant differences among the 36 *Brassica napus* genotypes for all the traits studied, indicating the presence of considerable genetic variability within the experimental material (Table 2). The existence of significant variation among genotypes suggests ample scope for selection and genetic improvement of important agronomic traits. Such variability is essential for the successful identification of superior genotypes and the development of improved cultivars through breeding programmes. The observed differences among genotypes may be attributed to their diverse genetic backgrounds, thereby providing opportunities for effective selection and hybridization in future crop improvement efforts.

Genetic variability

The estimates of environmental coefficient of variation (ECV), genotypic coefficient of variation (GCV), phenotypic coefficient of variation (PCV), heritability, genetic advance (GA), and genetic advance as percent

of mean (GAM) for the studied traits are presented in Fig. 1. In all traits, PCV values were slightly higher than the corresponding GCV values, indicating a minor influence of environmental factors on trait expression and suggesting that most of the observed variation was genetically controlled. High GCV and PCV values were observed for the number of siliquae on the main raceme, number of siliquae per plant, length of the main raceme, and seed yield per plant, indicating the presence of substantial genetic variability for these traits in Fig. 1. Moderate variability was recorded for plant height, number of primary branches, average seed per siliqua, and 1000-seed weight, whereas days to 50% flowering and days to maturity exhibited comparatively low variability. High heritability coupled with high genetic advance as percent of mean was observed for length of main raceme, number of siliquae on the main raceme, number of siliquae per plant, average seeds per siliqua, and seed yield per plant in Fig. 1. This suggests the predominance of additive gene action and indicates that direct selection for these traits would be highly effective. Conversely, traits such as days to maturity and oil content exhibited relatively lower genetic advance despite moderate to high heritability, suggesting the involvement of non-additive gene effects and environmental influence in Fig. 1. Among all the traits, the highest genetic advance as percent of mean was recorded for the number of siliquae on the main raceme, followed by average seeds per siliqua, length of main raceme, 1000-seed weight, and seed yield per plant in Fig. 1. Therefore, these traits can be considered important selection criteria for improving seed yield and overall performance in *Brassica napus* breeding programmes.

Diversity analysis

Cluster analysis yielded eight phenotypic groups, with the first group containing 14 genotypes, the second group containing 10 genotypes, the third group containing 6 genotypes, the fourth group containing 2 genotypes, and the other groups, such as 5, 6, 7, and 8, containing 1

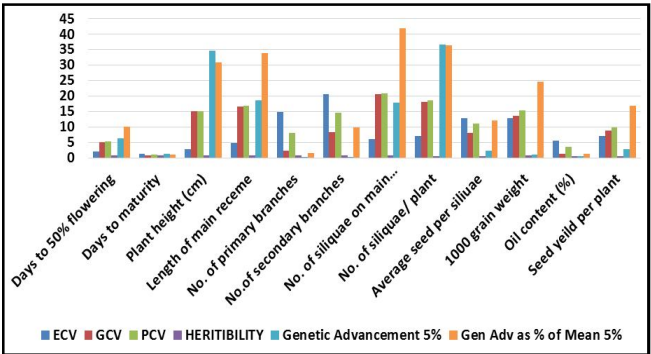


Fig. 1: Estimates of ECV, GCV, PCV, heritability, and genetic advance as a percentage of the mean for different quantitative traits in *Brassica napus* genotypes.

Table 3: Clustering pattern of 36 *Brassica napus* genotypes (including checks) for yield and its attributing traits.

Clus- ters	No. of genotypes	Genotypes
I	14	EC-338961, IC-399682, IC-491586, EC-657053, EC-338979, EC-358987, EC-657049, EC-400804, EC-414279, EC-358975, EC-386738, EC-338978, EC-414271, EC-414271
II	10	EC-339016, IC-399681, BM-91, EC-389916-, EC-657036, IC-339862, EC-338995, EC-338968, IC-399811, EC-339000
III	6	IC-278080, EC-338977, EC-339015, EC-414273, EC-338981, EC-414276
IV	2	IC-279074, EC-338993
V	1	EC-339001
VI	1	EC-338985
vii	1	EC-338994
viii	1	EC-338999

genotype each (Fig. 2 and Table 3). The spatial distance between two clusters indicates the degree of genetic divergence; a longer distance indicates greater divergence. As a result, genotypes within the same cluster exhibit less variation than genotypes from separate clusters (Table 4 and Fig. 3). Cluster 3 had the highest intra-cluster distance (8.34), followed by Cluster 2 (7.79), Cluster 1 (7.34), and Cluster 4 (4.48). Clusters 5, 6, 7, and 8 had the shortest intra-cluster distance (0.01, 0.00, 0.02, 0.00). It was found that the genotypes in Cluster 3 were the most genetically diverse, exhibiting a wide range of variation among them. In contrast, the genotypes in Clusters 5, 6, 7, and 8 were all genetically homogeneous. The cluster means analysis in Table 5 revealed that the genotypes have high genetic variability. Early flowering and maturity, tall plant stature, longer main raceme, more

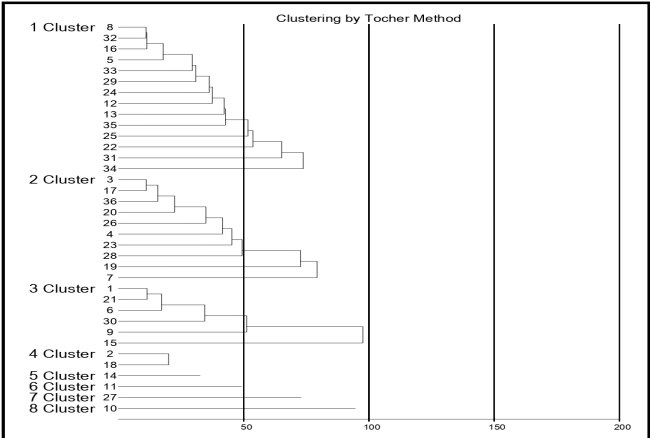


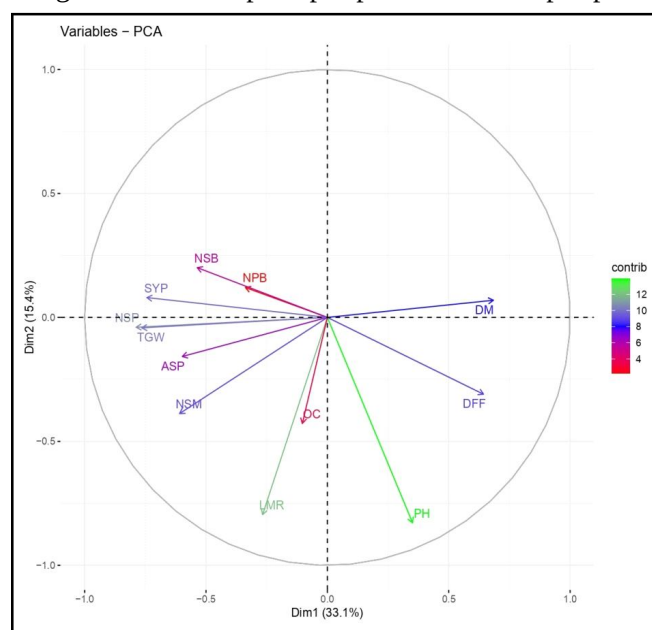
Fig. 2: Dendrogram showing the relationship between 36 genotypes, including checks of *Brassica napus* genotypes.

Table 4: Intra and inter-cluster distances (D^2) among eight clusters in thirty-six *Brassica napus* genotypes.

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Cluster 1	7.34	12.72	13.54	10.91	10.53	11.72	16.78	14.66
Cluster 2		7.79	14.72	13.91	18.38	19.54	11.84	23.97
Cluster 3			8.34	18.85	16.41	19.83	21.88	21.81
Cluster 4				4.48	14.37	14.1	14.06	15.89
Cluster 5					0.01	7.17	19.99	11.49
Cluster 6						0.00	20.88	10.02
Cluster 7							0.02	26.01
Cluster 8								0.00

primary and secondary branches, higher number of siliquae on main raceme and per plant, larger average seed per siliqua, higher 1000-grain weight, superior oil content, and increased seed yield per plant were identified as potential donors. These clusters can be useful for choosing parents in breeding programs intended to boost yield and contribute qualities in rapeseed.

A similar finding was reported by Yadav *et al.*, (2022). The cluster mean analysis indicated significant genetic variability among genotypes, with distinct clusters identified as potential sources of critical agronomic features. Groups 3 and 8 proved to be the best donor clusters for early flowering and maturity, respectively. Group 8 also showed potential for large plant stature; however, Group 5 had a longer main raceme length. Groups 7 and 3 had the highest cluster means for the number of primary and secondary branches, respectively. Srivastav *et al.*, (2002) found a similar conclusion that Group 5 and Group 3 had the most siliquae on the main raceme and per plant. It may be determined that rapeseed seed yield can be increased by selecting an ideotype with a higher number of pods per plant and seeds per plant.

**Fig. 3:** Variable loading for the first four principal components.

Choudhary and Joshi (2001) used cluster analysis to analyze genetic diversity across *Brassica* genotypes.

Principal Component Analysis

Principal component analysis (PCA) was performed to assess the pattern of genetic diversity among the 36 *Brassica napus* genotypes based on twelve quantitative traits. Among the twelve principal components extracted, the first four components had eigenvalues greater than unity and were therefore considered significant for interpretation (Table 6). These four principal components together explained 71.40% of the total phenotypic variation present among the genotypes. The first principal component (PC1) accounted for 33.10% of the total variation, followed by PC2 (15.40%), PC3 (14.60%), and PC4 (8.30%) (Table 6; Fig. 3). The substantial proportion of variability explained by these components indicates that they adequately summarize the genetic divergence among the evaluated genotypes. According to the criterion proposed by Johnson and Wichern (1988), traits with higher loading values contribute more significantly to the respective principal components. The first principal component (PC1) was mainly associated with length of

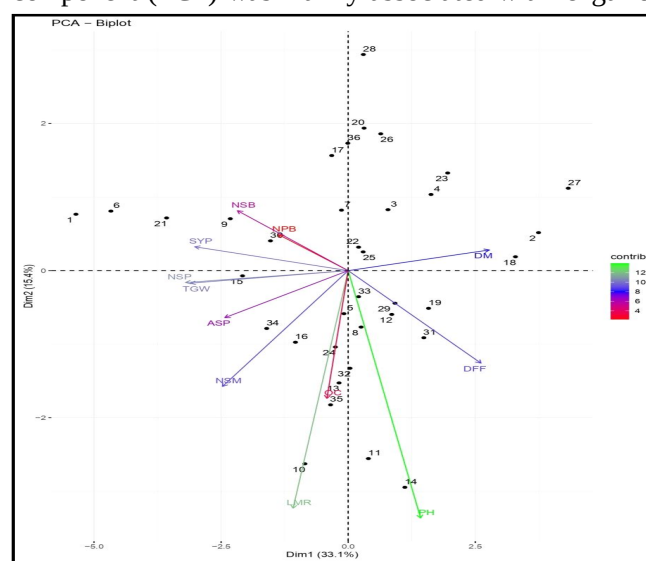
**Fig. 4:** PCA Biplot of 36 genotypes, including three checks of *Brassica napus* genotypes for yield and yield-related traits.

Table 5: Cluster mean for different characters in *Brassica napus* genotypes.

Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Days to 50% flowering	63.4	63.87	59.61	62.33	70	70.33	71	60.67
Days to maturity	123.02	123.7	120.83	123.5	123	123.67	124.33	120.67
Plant height (cm)	120.92	99.81	93.35	127.4	129.85	145.98	108.44	156.11
Length of main raceme (cm)	59.59	44.48	57.05	54.16	70.12	63.03	39.22	64.06
No. of primary branches	4.56	4.44	4.77	4.72	4	4.22	4.89	4.44
No. of secondary branches	2.77	2.87	3.41	3.33	2.67	3.22	2.66	3.22
No. of siliquae on main raceme	42.19	35.4	51.89	33.94	58.33	49.33	38.33	54.22
No. of siliquae per plant	97.53	95.55	128.42	84.17	95.44	115.89	52.89	102.11
Average seed per siliquae	19.95	19.63	20.07	13.78	17.44	19.89	14.55	19.78
1000 seed weight (g)	4.45	4.43	4.99	2.4	4.15	4.47	2.63	4.08
Oil content (%)	39.56	38.83	39.33	39	40.38	39.74	39.39	40.25
Seed yield per plant (g)	16.5	16.32	17.97	11.75	15.73	16.21	15.47	15.71

main raceme (LMR), number of secondary branches (NSB), number of siliquae on the main raceme (NSM), number of siliquae per plant (NSP), average seeds per siliqua (ASP), 1000-seed weight (TGW), seed yield per plant (SYP), and oil content (OC). Therefore, PC1 primarily represented yield-contributing and productivity-related traits. The second principal component (PC2) was largely influenced by length of main raceme, oil content, number of primary branches, and seed yield per plant. The third principal component (PC3) was predominantly associated with plant height, number of primary branches, number of secondary branches, and length of main raceme, whereas the fourth principal component (PC4) was mainly influenced by days to maturity, oil content, number of primary branches, and seed yield per plant.

The loading values further revealed that in PC1, positive contributions were recorded for length of main raceme (0.134), number of primary branches (0.169), number of secondary branches (0.269), number of siliquae on the main raceme (0.305), number of siliquae per plant (0.395), average seeds per siliqua (0.300), 1000-seed weight (0.383), oil content (0.052), and seed yield per plant (0.373). In contrast, days to 50% flowering (-0.323), days to maturity (-0.343), and plant height (-0.176) exhibited negative loadings, indicating an inverse association with the traits contributing positively to PC1. In PC2, positive loadings were observed for days to 50% flowering (0.228), length of main raceme (0.586), number of siliquae per plant (0.029), average seeds per siliqua (0.116), 1000-seed weight (0.031), and oil content (0.315), while several other traits contributed negatively. Similarly, PC3 showed positive loadings for plant height (0.131), length of main raceme (0.193), number of primary branches (0.373), number of secondary branches (0.467), number of siliquae on the main raceme (0.137), and number of siliquae per plant (0.032). In PC4, positive contributions were recorded for days to maturity (0.270),

number of primary branches (0.480), number of secondary branches (0.110), oil content (0.778), and seed yield per plant (0.101), whereas the remaining traits exhibited negative loadings. The PCA biplot based on PC1 and PC2 (Fig. 4) illustrated the relative contribution of different traits to the overall genetic variability among the genotypes. Traits such as plant height, length of main raceme, oil content, number of siliquae on the main raceme, average seeds per siliqua, 1000-seed weight, number of siliquae per plant, and seed yield per plant were positioned farther from the origin, indicating their greater contribution to genotype differentiation. These traits were identified as the major determinants of genetic diversity in the present germplasm collection. The present findings are in agreement with earlier reports by Yu *et al.* (2007), Khan *et al.*, (2014), Neeru *et al.*, (2015), Gupta

Table 6: Principal component analysis for 12 quantitative traits in *Brassica napus* germplasm.

Traits	PC1	PC2	PC3	PC4
Days to 50% flowering	-0.323	0.228	-0.229	-0.027
Days to maturity	-0.343	-0.051	-0.242	0.27
Plant height	-0.176	0.61	0.131	-0.015
Length of main raceme	0.134	0.586	0.193	-0.143
No. of primary branches	0.169	-0.089	0.373	0.48
No. of secondary branches	0.269	-0.148	0.467	0.11
No. of siliquae on main raceme	0.305	0.286	0.137	-0.203
No. of siliquae/ plant	0.395	0.029	0.032	-0.003
Average seed per siliquae	0.3	0.116	-0.431	-0.08
1000 seed weight (g)	0.383	0.031	-0.384	-0.005
Oil content %	0.052	0.315	-0.151	0.778
Seed yield per plant (g)	0.373	-0.059	-0.318	0.101
Eigenvalue	3.976	1.843	1.750	1.002
Proportion %	33.100	15.400	14.600	8.300
Cumulative %	33.100	48.500	63.100	71.400

et al., (2019), Ilyas *et al.*, (2018), and Aslam *et al.*, (2025), who also reported that the first few principal components accounted for a major proportion of the total variation in Brassica species. Therefore, the genotypes identified through the first four principal components may serve as valuable genetic resources for parent selection and future breeding programmes aimed at improving seed yield, oil content, and related agronomic traits in *Brassica napus*.

Discussion

The highly significant differences detected among the 36 *Brassica napus* genotypes for all traits confirm the presence of substantial exploitable genetic variability in the panel, in agreement with earlier reports on rapeseed–mustard germplasm. This broad variation provides a sound basis for effective selection and recombination, similar to the patterns documented by Rameeh *et al.*, (2015) for quantitative traits in *B. napus*.

The present study revealed substantial genetic variability among the *Brassica napus* genotypes for the traits evaluated. The slightly higher phenotypic coefficient of variation (PCV) values compared with the corresponding genotypic coefficient of variation (GCV) values for all traits indicated a relatively minor influence of environmental factors on trait expression, suggesting that most of the observed variation was genetically controlled. High GCV and PCV estimates for the number of siliquae on the main raceme, number of siliquae per plant, length of the main raceme, and seed yield per plant demonstrated considerable genetic variability and substantial scope for effective selection. Similar findings have been reported in earlier studies on Brassica germplasm. High heritability coupled with high genetic advance as a percentage of the mean for plant height, length of the main raceme, number of siliquae on the main raceme, number of siliquae per plant, and seed yield per plant indicated the predominance of additive gene action, making these traits suitable targets for direct phenotypic selection. Under such genetic control, the observed variation largely reflects heritable differences rather than environmental effects (Yu *et al.*, 2009). In contrast, days to maturity and oil content exhibited moderate to high heritability but relatively low genetic advance, suggesting the involvement of non-additive gene effects, a narrower genetic base, or greater environmental influence on trait expression. Similar observations have been reported in Brassica by Aslam *et al.*, (2025). Therefore, siliquae-related traits and seed yield may serve as reliable selection criteria for improving productivity in *Brassica napus* breeding programmes.

The divergence analysis further clarified the structure

of variability and highlighted specific groups as promising donor pools. The formation of eight distinct clusters, with the largest groups harbouring substantial intra-cluster variability and clusters 5–8 being essentially homogeneous, indicates that both wide and narrow genetic bases coexist within the evaluated material (Yadav *et al.*, 2025). The high intra-cluster distances in clusters 3, 2, and 1 emphasise their internal heterogeneity, whereas the zero intra-cluster distances in clusters 5–8 imply that these groups may represent nearly uniform lines or closely related genotypes, a pattern consistent with earlier reports of discrete diversity blocks in Brassica germplasm (Choudhary & Joshi, 2001).

Cluster mean performance identified specific clusters as potential donors for key agronomic attributes, which is crucial for designing crossing blocks and ideotypes. Clusters 3 and 8 were particularly important for early flowering and early maturity, respectively, while cluster 8 also contributed to tall stature and cluster 5 provided longer main racemes. The highest cluster means for primary and secondary branches in clusters 7 and 3, along with elevated siliqua number and seed yield, echo previous findings that donor clusters with enhanced branching and siliqua production can substantially improve yield potential when used as parents (Yadav *et al.*, 2022). These results support the strategy of selecting parents from widely divergent and agronomically superior clusters to maximize heterosis and favourable transgressive segregation, especially for pods per plant, seeds per siliqua, and seed yield per plant, which have repeatedly been associated with enhanced yield in rapeseed (Khan *et al.*, 2018).

The principal component analysis complemented the cluster analysis by partitioning the multivariate variability into a few informative dimensions. The first four principal components with eigenvalues greater than one together explained over 70% of the total variation, indicating that a relatively small number of underlying trait complexes capture most of the genetic divergence among genotypes (Yu *et al.*, 2007). PC1 was heavily weighted by length of main raceme, number of siliquae on the main raceme, number of siliquae per plant, average seed per siliqua, 1000-seed weight, oil content, and seed yield per plant, suggesting that it represents a yield–yield-component axis where raceme architecture and seed size jointly drive productivity, similar to the primary PCs reported for *B. napus* and *B. juncea* in earlier multivariate studies (Chakraborty *et al.*, 2021). PC2 was dominated by primary and secondary branches and seed yield per plant, emphasizing the role of branching pattern and sink capacity as a second major dimension of variation. PC3 and PC4 were more closely associated with phenology and plant

stature, with days to 50% flowering and plant height loading strongly on PC3, and days to maturity, branching traits, and oil content contributing positively to PC4. Such separation of phenology, architectural, and yield-component traits into distinct components agrees with previous PCA-based characterizations of Brassica germplasm, which also observed that the first few PCs capture combinations of flowering time, plant height, and silique traits responsible for most of the observed diversity (Aslam *et al.*, 2025). The positive and negative loadings within PCs offer practical guidance for simultaneous improvement or avoidance of undesirable correlations. In PC1, the negative loadings of days to 50% flowering, days to maturity, and plant height contrasted with positive loadings for yield-related traits, implying that early and moderately short genotypes can still achieve high seed yield when associated with longer racemes, more siliques, and heavier seeds. Conversely, in PC2 and PC3, the positive contributions of branching traits and raceme length indicate that selection for increased branching must be carefully balanced with raceme architecture and phenology to avoid excessive vegetative growth at the expense of harvest index (Chakraborty *et al.*, 2021). The strong contributions of oil content and seed yield in PC4 further suggest that genotypes combining high oil concentration with acceptable yield can be identified by targeting favorable scores on this component. The biplot of PC1 and PC2, integrating trait loadings and genotype positions, underscores that plant height, length of main raceme, number of siliques on the main raceme, number of siliques per plant, 1000-seed weight, oil content, and seed yield per plant are the principal contributors to genetic differentiation among the *B. napus* genotypes evaluated. This pattern reinforces the evidence from variance and cluster analyses that these traits are key drivers of adaptation and yield potential in the current germplasm set and should be prioritized in selection indices. Similar multivariate biplots in rapeseed and mustard have also highlighted raceme length, silique density, and seed size as central traits structuring genetic diversity, confirming the robustness of these parameters as descriptors in Brassica improvement programs (Khan *et al.*, 2018; Neeru *et al.*, 2015).

From a breeding perspective, the convergence of high heritability with high genetic advance for yield-related traits, clear inter-cluster divergence, and the concentration of key yield components in the first principal components provides a coherent framework for parent selection and early-generation screening. Crosses between high-yielding, well-diverged clusters such as 3, 5, 7, and 8 are likely to generate wide recombination for raceme

architecture, silique number, and branching, with sufficient additive variance to respond to selection. Emphasis on genotypes with favorable PC1 and PC2 scores—combining early to medium maturity, optimal plant height, longer main racemes, more siliques per plant, and high 1000-seed weight—should accelerate the development of ideotypes with improved seed yield and oil content. Future work integrating these multivariate patterns with molecular marker data could further refine the identification of heterotic pools and facilitate marker-assisted selection for complex yield and quality traits in *B. napus*.

Conclusion

Several genotypes and trait combinations emerged as particularly promising for use as parents in rapeseed improvement, based on both cluster and principal component analyses. Genotypes grouped in Cluster 3 and Cluster 8 were especially valuable donors: Cluster 3 showed the highest intra cluster distance (8.34), indicating wide internal diversity, and contained genotypes with superior branching and silique traits, while Cluster 8 combined early maturity with tall plant stature and high mean values for key yield components (Table 4). Cluster means revealed genotypes with early flowering and maturity, tall plants, longer main racemes, more primary (NPB) and secondary branches (NSB), more siliques on the main raceme (NSM) and per plant (NSP), higher average seeds per silique (ASP), greater 1000 seed weight (TGW), higher oil content (OC), and increased seed yield per plant (SYP), identifying them as potential donors for multiple desirable traits. In the principal component analysis, the first four PCs explained 71.4% of total variation, with PC1 heavily weighted by length of main raceme (0.134), number of primary branches (0.169), number of secondary branches (0.269), siliques on main raceme (0.305), siliques per plant (0.395), average seed per silique (0.300), 1000 seed weight (0.383), oil content (0.052), and seed yield per plant (0.373), indicating that genotypes with high scores for these traits in PC1 represent the best candidates for simultaneous improvement of yield and quality. Similarly, PC2 emphasized raceme length (0.586), oil content (0.315), and earliness, while PC3 and PC4 captured variation in plant height, branching, days to flowering, and maturity, helping pinpoint genotypes that combine earliness with robust plant architecture. Overall, genotypes from Clusters 3, 7, and 8, with high cluster means and strong positive loadings for these traits, can be considered the most promising materials for parent selection aimed at improving yield, seed size, silique number, and oil content in *Brassica napus*.

Author Contributions

Parul Pandey: Conceptualization, Methodology, Investigation, Data curation, Writing original draft. Shweta: Conceptualization, Supervision, Project administration, Writing review & editing. Majiga Revanth Kumar: Methodology, Formal analysis, Validation, Writing review & editing. Sandeep Varma: Investigation, Data curation, Visualization, Writing review & editing. Nitin Kumar: Formal analysis, Data interpretation, Visualization, Writing – original draft, Writing review & editing. Nikhil Kumar: Investigation, Data curation, Validation, Writing review & editing. Chandragupt Maurya: Resources, Investigation, Data curation, Writing review & editing. Amit Kumar: Supervision, Resources, Writing review & editing. Shivam Tripathi: Methodology, Validation, Formal analysis, Writing review & editing. All authors read and approved the final version of the manuscript.

References

- Ali, N., Khan N.U., Farhatullah R.M., Bibi Z., Gul S., Ali S. and Naveed K. (2018). Genetic diversity in indigenous landraces of *Brassica napus* based on morphological and biochemical characteristics using multivariate techniques. *Int. J. Agric. Biol.*, **20**(2), 277-287.
- Aslam, M., Khan T.A., Ali Z. and Hussain N. (2025). Multivariate analysis of oil and yield traits in rapeseed. *International Journal of Oilseed Crops Research*, **12**(1), 23–30.
- Chakraborty, S., Kumar A., Kishore C., Singh R.S., Kumar A., Kumar A. and Kumar M. (2021). Multivariate based diversity analysis in Indian mustard (*Brassica juncea* L.) genotypes by principle component analysis. In *Biological Forum—An International Journal*, **13**(3), 680-685.
- Choudhary, B.R. and Joshi P. (2001). Genetic diversity in advanced derivatives of *Brassica* interspecific hybrids. *Euphytica*, **121**(1), 1-7.
- Ciancaleoni, S., Chiarenza G.L., Raggi L., Branca F. and Negri V. (2014). Diversity characterisation of broccoli (*Brassica oleracea* L. var. italica Plenck) landraces for their on-farm (in situ) safeguard and use in breeding programs. *Genetic resources and crop evolution*, **61**(2), 451-464.
- Diepenbrock, W. (2000). Yield analysis of winter oilseed rape (*Brassica napus* L.): a review. *Field crops research*, **67**(1), 35-49.
- Labana, K.S. and Banga S.S. (1993). *Oilseed Brassicas: Advances in Research and Future Implications*. Kalyani Publishers.
- Marjanovic-Jeromela A., Marinkovic, R., Ivanovska S., Jankulovska M., Mijic A. and Hristov N. (2011). Variability of yield determining components in winter rapeseed (*Brassica napus* L.) and their correlation with seed yield. *Genetika-Belgrade*, **43**(1), 51-66.
- Parvin, E., Mahmud F., Bhuiyan S.R. and Haque M. (2019). Multivariate analysis of genetic variation in rapeseed (*Brassica Napus* L.). *Agriculture and Food Sciences Research*, **6**(1), 1-8.
- Rameeh, V. (2015). Genetic variability and interrelationships among quantitative traits in rapeseed (*Brassica napus* L.) advanced lines. *Journal of Agricultural Sciences—Sri Lanka*, **10**(3).
- Rao, C.R. (1952). Advanced statistical methods in biometrical research. 1st edition. New York: John Wiley and Sons; 236-72.
- Shehzad, M.A., Khan M.I., Iqbal M. and Shah S.H. (2015). Genetic diversity assessment in *Brassica napus* using agro-morphological traits. *Journal of Agricultural Research*, **53**(2), 179–188.
- Yadav, A., Yadav S.K., Choudhary R., Kumar S.M., Anumantharaj A., Bharadwaj R. and Yadav S. (2025). Insights from multivariate analysis of seed vigor traits in rapeseed and mustard (*Brassica* spp.) genotypes. *Turkish Journal of Agriculture and Forestry*, **49**(4), 651-660.
- Yu, S., Zhang F., Yu R., Zou Y., Qi J., Zhao X. and Li L. (2009). Genetic mapping and localization of a major QTL for seedling resistance to downy mildew in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Molecular Breeding*, **23**(4), 573-590.