RESEARCH ARTICLE

Integrative multivariate and clustering analysis of field pea in late sown condition

Aman Srivastava[1], Shiva Nath[1], Banshidhar[1], Priyanka Jaiswal[1], Prasenjit Paul[2], Sonali Srivastava[1], Dharmendra Kumar[1], Priya Raj[1], Vinay Kumar[1], Poonam Sharma[3] & Anamish Tyagi[1*]

[1]Department of Genetics and Plant Breeding, Acharya Narendra Deva University of Agriculture and Technology, Kumarganj, Ayodhya 224 229, Uttar Pradesh, India
[2]Department of Genetics and Plant Breeding, Dhamma Dipa International Buddhist University, Sabroom 799 143, South Tripura, India
[3]Department of Genetics and Plant Breeding, Chaudhary Sarwan Kumar Himachal Pradesh Krishi Vishvavidyalaya, Palampur 176 062 Himachal Pradesh, India

*Correspondence email - anamishtyagi1999@gmail.com

Cite this article: Aman S, Shiva N, Banshidhar, Priyanka J, Prasenjit P, Sonali S, Dharmendra K, Priya R, Vinay K, Poonam S, Anamish T. Integrative multivariate and clustering analysis of field pea in late sown condition. Plant Science Today (Early Access). https://doi.org/10.14719/pst.9339

Abstract

Field pea is an important rabi crop in India. This study evaluated 10 parent varieties and 45 half diallel crosses to understand their genetic diversity under late-sown conditions. Eleven different traits were measured and the D2 Mahalanobis clustering method was employed to analyse the data, revealing that the germplasm could be grouped into six distinct clusters. The largest genetic differences were observed between clusters IV and V. Analysis of traits revealed significant variations among the clusters, with some genotypes from clusters IV and V showing promise for breeding purposes. Principal component analysis identified four key components that explained 83.61 % of the total variation. The highest variation was explained by PC1 (39.43 %) with an eigen value of 4.3, highlighting important traits for future crop improvement efforts.

Keywords: d2 mahalonobis; field pea: genetic diversity; principal component

Introduction

Field peas are self-pollinating plants with chromosome number $2n = 14$ that belong to the genus and species Sativum, subfamily Papilionaceae and family Leguminosae (Fabaceae). Among the genus Pisum, species sativum subsp. Abyssinicum is considered a possible candidate as progenitor and resembles the cultivated form closely (1). Field pea is a succulent, semi-upright to erect annual herbaceous plant, that can reach heights of 30 to 200 cm and have a propensity to climb when support is available. The taproot system of plants has surface nodules. Usually compound, leaves have terminal, branching tendrils and one to three pairs of leaflets.

Nutritionally, pea is a rich source of protein, ranging from 21-25 % with higher concentrations of lysine and tryptophan amino acids (2-3) and have low level of cysteine and methionine amino acids (4). Ripe, mature, dried field pea seeds are used as flour, whole in the form of chhola chat, or as pulse (split seeds). They have nutritional value, but they can also increase soil fertility by forming a symbiotic relationship between nitrogen-fixing bacteria and plant root nodules. Therefore, cultivation of pulses elevates the productivity of soil in terms of the yield potential of subsequent crops by 20-40 % has been recorded (5).

The population of India is expected to exceed 1.68 billion people by 2030, according to the vision of IIPR, Kanpur. It is projected that the demand for pulses will reach 32 million tons in 2030, with a required annual growth rate of 4.2 %. To meet the challenges of an expanding population and boost availability per capita, a quantum increase in pulse production is necessary. Therefore, a high-yielding field pea variety with high-quality seed is required. The selection of genetically varied parental genotypes for the hybridization program is predicated on the idea that Crosses involving divergent parents offer a greater possibility of obtaining desirable segregants in the segregating generations. Several researchers addressed the need of diverse parent to obtain superior genotypes in the segregating generations (6-7). The use of advanced statistical methods in agricultural research has become increasingly important for understanding the genetic diversity of crops. One such method is Principal Component Analysis (PCA), which helps in visualizing the genetic distance among individuals by representing data in two or three dimensions. PCA simplifies complex data by breaking down the total variance into a smaller set of variables, thereby revealing the relationships between different parameters. The first principal component accounts for the most variability in the data, while the subsequent components capture the remaining variability. This technique has been effectively utilized in studies involving *Pisum sativum* L. (pea), as shown in the previous work (8-9).

Another valuable approach is cluster analysis, which groups individuals based on their similarities. This method creates clusters that exhibit high homogeneity within each group and significant differences between groups, making it easier to visualize genetic relationships. The distance-based methods used in cluster analysis involve a pair-wise distance matrix, resulting in graphical representations such as trees or dendrograms. This technique has been widely applied in crop species to analyze genetic diversity, including research on *Pisum sativum* L. conducted by former researchers (10).

Recognizing the importance of the pea crop, this study aims to assess the available germplasm to facilitate the development of high-yielding varieties of *Pisum sativum* L. Understanding genetic diversity is crucial for enhancing crop improvement efforts and employing techniques like PCA and Cluster Analysis will provide valuable insights into the genetic makeup of pea varieties.

Materials and Methods

Seed material and Location

The research trial was conducted over two successive years. Initially (2022-23), ten genetically diverse parents were selected, crosses were made among them and 45 F_1 were obtained through half-diallel mating strategy all grown in randomized block design (11) . In the next seasons (2023-24), all the 45 F_1 along with the ten parents were grown under late-sown condition at Genetics and Plant breeding farm of Acharya Narendra Deva University of Agriculture and Technology, Ayodhya, Uttar Pradesh, India situated at longitude 81.824688^0 with a latitude 26.541242^0 during the Rabi seasons. The detail of the parents and there F_1 progeny were given separately in Supplementary Table 1. Three replication blocks were created from the entire experimental site. Every plot has a single row of genotypes that are 5 meters long and spaced 30 by 10 cm apart.

Phenotypic estimation of quantitative traits

The present investigation involved the study of eleven quantitative traits which are used to assess the crop performance. The different parameter which were used under the study are as following: days to maturity (DM), days to flowering (DF), plant height (PH), number of primary branches per plant (NPBPP), number of secondary branches/plant (NSBPP), pods per plant, seeds per pod, 100-seed weight, seed yield/plant (SY), harvest index (HI) and biological yield/plant (BY).

Statistical analysis

The mean data of each character was collected by choosing five plants randomly. Agrianalyze software is used for measuring genetic divergence through principal component analysis (based on R and Python programming). Mean values were subjected to cluster analysis by using D2 Mahalanobis clustering method. The D^2 distance between the

i^{th} and j^{th} genotypes for K characters was computed as:

$$D^2ij = \sum_{t=1}^{K}(Yij - Yjt)^2$$

Where, t =1

Results and Discussion

Our analysis of 55 field pea genotypes showed significant genetic differences between all traits. We observed large distances between different clusters compared to distances within the same cluster, indicating that the clusters are distinct. Notably, the greatest genetic separation was found between Cluster I and Cluster V, suggesting these clusters have very different genetic backgrounds. In contrast, Clusters I and II were found to be genetically similar. When we examined the average traits of the clusters, Clusters V and VI stand out for their productivity. Cluster VI had the highest average seed yield (SY) and biological yield (BY), while Cluster V had the best harvest index (HI) and longest pod length (PL). This finding is like previous studies, which identified clusters with high SY and BY, as well as previous researchers, who reported clusters that excelled in yield-related traits (12). In our study, Clusters V and VI produced a high number of pods and seeds, with Cluster VI having the top BY at 38.73 g/plant and SY at 14.68 g/plant. Other traits also varied significantly among the clusters (13). For example, Cluster II had the latest flowering and maturity in one comparable study, while our data showed the earliest flowering and maturity occurred in a different cluster. Also, plant height and branching were different; Cluster IV, according to previous reports, had the tallest plants and the most branches, indicating strong vegetative growth (14). Some clusters had many small seeds, whereas others had fewer but larger seeds. This means each cluster had its unique combination of traits across all 13 characteristics, such as days to flowering, days to maturity, plant height and seed yield. These results have important implications for breeding. The large genetic distances suggest that crossing genotypes from very different clusters will yield high diversity and potential hybrid vigour. For instance, crossing a Cluster .VI line, which is high in SY and BY, with a Cluster V line, known for its high HI, could combine their strengths. Other studies, also recommend crossing plants from the most divergent clusters for better offspring (14-16).

Furthermore, the composition of traits in each cluster indicates specific goals for selection. Breeders aiming for high yield should focus on Clusters V and VI, as these already show superior SY, BY and HI. If early maturity is desired, genotypes from clusters with shorter flowering and maturity times can be crossed with high-yield clusters. Clusters characterised by greater branching or pod numbers, such as the taller Cluster IV, could also provide useful traits. Cluster details are given in Table 1-2. Fig. 1 showing relationship between parents and half diallel crosses in field pea.

Principal component analysis (PCA)

PCA is a statistical method of multivariate analysis that reduces the set of many variables to a set of small linearly uncorrelated variables, which can explain most of the variation present in the original variables (17). According to the results of PCA, the yield component traits of field pea germplasm showed a maximum variability of approximately 81.63 %, with only the first four principal components (PCs) exhibiting more than 1.00 eigenvalue. The Principal Component Analysis was done using 11 traits. The singular value decomposition approach was used for analysis. The highest variation was explained by PC1

Table 1. Estimates of Intra and inter cluster distance for 6 cluster in field pea

	intracluster distance	Intracluster distance					
		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	cluster 6
Cluster 1	24.65	0					
Cluster 2	0.00	42.67	0				
Cluster 3	21.79	47.64	27.51	0			
Cluster 4	27.17	163.84	79.22	72.30	0		
Cluster 5	23.08	65.22	146.41	148.91	340.77	0	
Cluster 6	0.00	143.99	90.83	70.03	52.92	291.83	0

Table 2. Cluster mean values of 11 traits of field pea

	NOG	d50	dtm	Ph (cm)	PBPP	NSBPP	pod/plant	seed/pod	100 tw	hi	by	sy
Cluster 1	26	52.90	96.19	83.87	2.53	4.51	27.09	5.06	26.44	34.76	103.99	36.07
Cluster 2	1	51.00	98.00	68.38	2.93	3.20	22.47	4.27	26.24	27.44	91.55	25.12
Cluster 3	7	53.10	97.57	90.25	2.54	3.97	24.30	5.36	24.29	37.92	82.29	31.04
Cluster 4	8	57.58	103.17	92.67	2.70	2.33	20.27	4.83	22.11	40.81	56.93	22.94
Cluster 5	12	54.61	95.78	95.78	2.43	4.72	31.56	5.05	26.35	32.94	127.16	41.80
Cluster 6	1	61.33	110.67	91.40	3.00	1.33	23.03	6.07	23.82	57.64	58.26	33.39

NOG- number of genotypes; D50- days to 50 % flowering; dtm- days to maturity; ph- plant height; PBPP- primary branches plant[-1]; NSBPP- number of secondary branches per plant; 100tw- 100 test weight; hi- harvest index ; by- biological yield; sy- seed yield
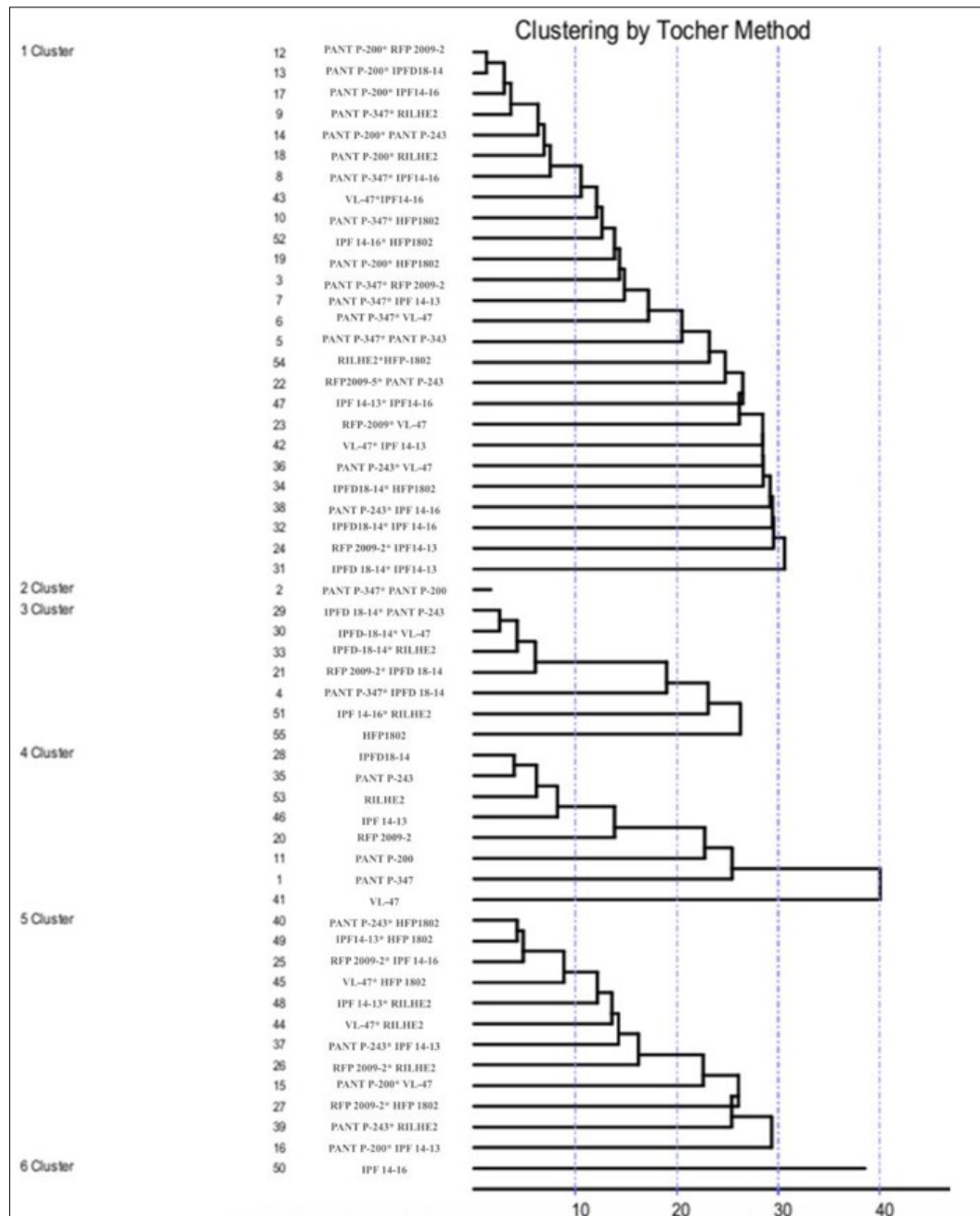
Fig. 1. Dendrogram showing relationship between parents and half diallel crosses in field pea.

(39.43 %) with an eigenvalue of 4.3 (Table 3, Fig. 2-3). The principal components PC2, PC3 and PC4 explained variation of 18.83 %, 12.77 % and 10.6 %, respectively. Similar findings were also found in previous studies (18-19). Eigenvalues help to decide how many variables to retain. The sum of the eigenvalues is generally equal to the number of variables (19-20).

Table 3. Eigen value and variability of different characters in pea

Principal Component	Eigen value	Proportion of Variance	Commulative var.
PC1	4.3	39.431	39.431
PC2	2.1	18.83	58.261
PC3	1.4	12.771	71.032
PC4	1.2	10.602	81.634

Inference from correlation between Principal Components and Variables

The first principal component (PC1) shows a positive correlation with several yield-related variables, including seed yield (0.4573), biomass (0.4378), the number of secondary branches per plant (0.3841), 100 seed weight (0.3823), pod count per plant (0.3665), primary branches per plant (0.1514), seeds per pod (0.1109), harvest index (0.1108) and plant height (0.1017). In contrast, PC1 exhibits a negative correlation with the number of days to 50 % flowering (-0.1932) and days to maturity (-0.2757). This indicates that as the values of the yield-related traits increase, the days to flowering and maturity tend to decrease, as illustrated in Table 4 and Fig. 4.

The principal component PC2 had positive correlation with variables number of secondary branches plant^{-1} (0.2833) by (0.1552), pod/plant (0.0925) and 100 test weight (0.034), while negative correlation with variables days to 50 % flowering (-0.0276), seed yield (-0.092), days to maturity (-0.1057), seed/pod (-0.2943), Plant height (-0.3475), primary branches plant^{-1} (-0.5264) and harvest index (-0.6173).

The principal component PC3 had positive correlation with variables 100 test weight (0.2053), harvest index (0.0945), seed/pod (0.0507) and primary branches plant^{-1} (0.0268), while negative correlation with variables number of secondary branches plant^{-1} (-0.0318), seed yield (-0.174), plant height (-0.2216), biological yield (-0.2297), pod/plant (-0.3857), days to maturity (-0.466) and days to 50 % flowering (-0.6682).
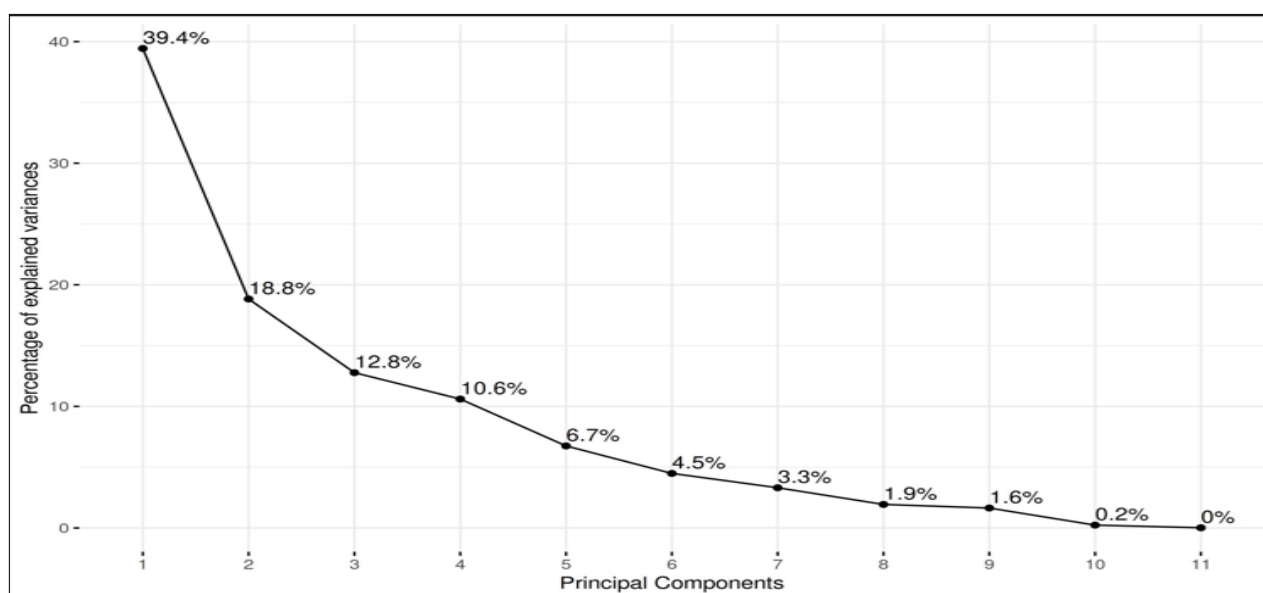
Fig. 2. Scree plot showing variation of commulative variance for different principal components.
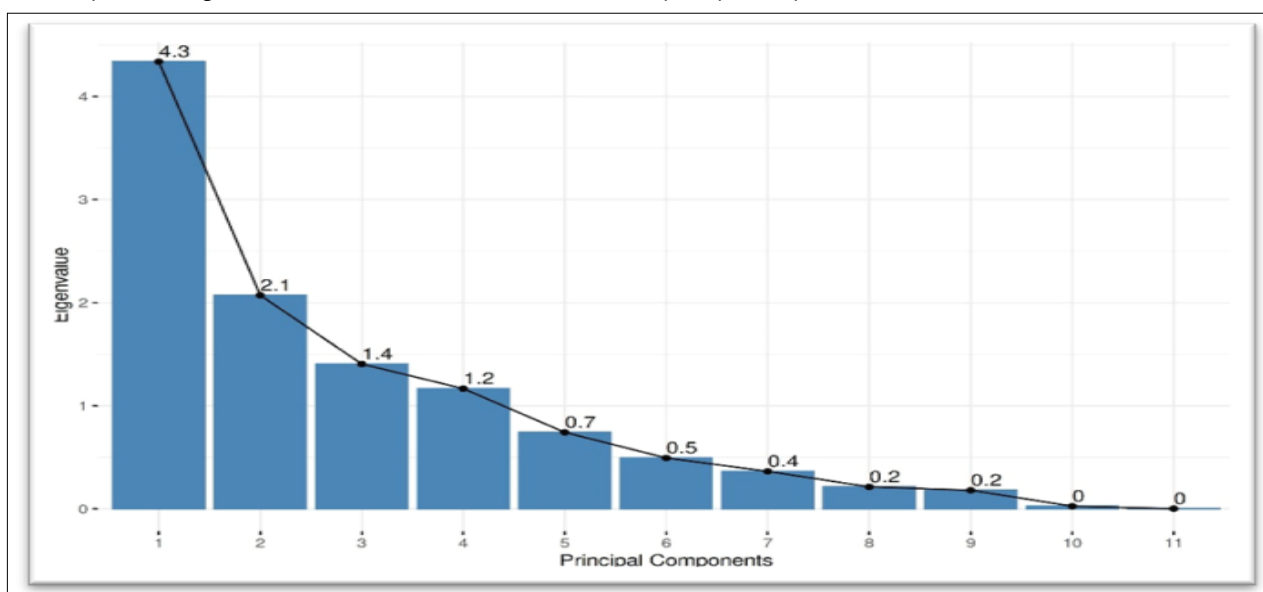

Fig. 3. Scree plot showing variation of commulative variance for different principal components.

Table 4. Correlation of different variables with principal components

Traits	PC1	PC2	PC3	PC4
Days to 50 % flowering	-0.1932	-0.0276	-0.6682	0.1172
Days to maturity	-0.2757	-0.1057	-0.466	-0.1848
Plant height	0.1017	-0.3475	-0.2216	0.3315
Primary branches plant⁻¹	0.1514	-0.5264	0.0268	-0.41
Secondary branches plant⁻¹	0.3841	0.2833	-0.0318	0.142
Pod/plant	0.3665	0.0925	-0.3857	-0.2702
Seed/pod	0.1109	-0.2943	0.0507	0.7331
100 test weight	0.3823	0.034	0.2053	-0.1494
Harvest index	0.1108	-0.6173	0.0945	-0.1266
Biological yield	0.4378	0.1552	-0.2297	0.0716
Seed yield	0.4573	-0.092	-0.174	0.0066
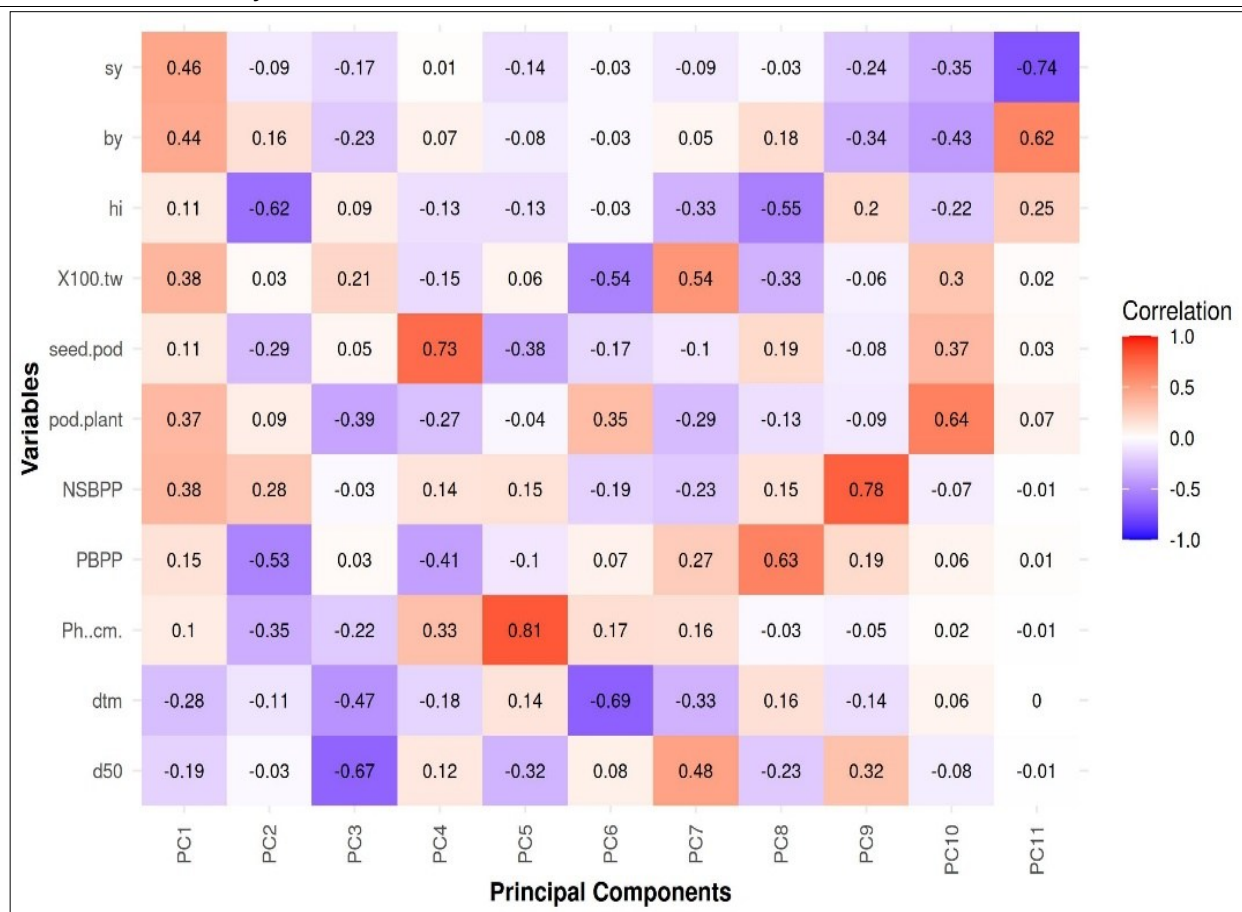
Fig. 4. Correlation of different variables with principal components Heat map.

The principal component PC4 showed a positive correlation with the variables seed/pod (0.7331), plant height (0.3315), number of secondary branches per plant (0.142), days to 50 % flowering (0.1172), biological yield (0.0716) and seed yield (0.0066), while negative correlation with the variables harvest index (-0.1266), 100 test weight (-0.1494), days to maturity (-0.1848), pod/plant (-0.2702) and primary branches per plant (-0.41).

Conclusion

Crosses between cluster members with high cluster means for significant characters and high inter-cluster distances are expected to yield more desirable recombinants in segregating generations. Several genotypes from clusters V and IV could be analyzed to isolate desirable segregants. Based on the results of PCA analysis, there are significant variations among the traits, with four main PCs describing 81.63 % of the total variation. It can be concluded that progenies resulting from diverse crosses should exhibit a wide range of genetic variability in the advanced generation and a greater possibility of separating transgressive segregants from the background population.

Authors' contributions

AS carried out the trial and drafted the manuscript. SN contributed to data collection. BD helped in data interpretation. PJ participated in the study design and assisted with statistical analysis. PP and AT contributed to experimental setup and data acquisition. SS helped with data tabulation. DK contributed to sequence validation and reviewed the manuscript. PR participated in the design of the study and contributed to manuscript revision. VK conceived of the study, participated in its design and coordination and supervised the overall project. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: The authors affirm that there are no financial, personal, or professional conflicts of interest that could have influenced the outcome or interpretation of this study. This research was conducted with integrity, transparency and a shared commitment to advancing scientific knowledge in the field of plant science. All authors have reviewed the manuscript and agree with its content. Authors do not have any conflict of interests to declare.

Ethical issues: None

References

1. Jing R, Vershinin A, Jacek G, Paul S, Petr S, David FM, et al. The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon-based insertion polymorphism (RBIP) marker analysis. BMC Evol Biol. 2010;10(1):44. https://doi.org/10.1186/1471-2148-10-44

2. Hagenblad J, Bostrom E, Nygards L, Leino MW. Genetic diversity in local cultivars of garden pea (*Pisum sativum* L.) conserved on farm and in historical collections. Genet Resour Crop Evol. 2014;61(2):413-22. https://doi.org/10.1007/s10722-013-0046-5

3. Egamberdieva D, Eshboev F, Jabbarov Z, Botirova D, Nazarov K, Khamidov M, et al. Challenges and prospects for grain legume production in Uzbekistan: a review. Environ Sustain. 2025;1-11. https://doi.org/10.1007/s42398-025-00346-y

4. Ceyhan E, Avci MA. Combining ability and heterosis for grain yield and some yield components in pea (*Pisum sativum* L.). Pak J Biol Sci. 2005;8(10):1447-52. https://doi.org/10.3923/pjbs.2005.1447.1452

5. Ouafi L, Alane F, Rahal-Bouziane H, Abdelguerfi A. Agro-morphological diversity within field pea (*Pisum sativum* L.) genotypes. Afr J Agric Res. 2016;11(40):4039-47. https://doi.org/10.5897/AJAR2016.11454

6. Kumar R, Kumar M. Estimation of genetic divergence in garden pea (*Pisum sativum* var. *hortense* L.) germplasm to facilitate the selection of potential parents for hybridization programme. Legume Res. 2016;39(5):709-12. https://doi.org/10.18805/lr.v0iOF.6851

7. Arif U, Ali S, Alharthi B, Pasha MFK, Ahmad HM, Ahmed MJ, et al. Genetic diversity analysis of pea (*Pisum sativum* L.) landraces by utilizing simple sequence repeat markers. Genet Resour Crop Evol. 2024;1-7. https://doi.org/10.1007/s10722-023-01796-1

8. Gonçalves L, Rubiales D, Lourenço M, Patto MCV. Exploring grass pea (*Lathyrus sativus* L.) genetic diversity in Mediterranean changing climate conditions. Eur J Agron. 2024;156:127142. https://doi.org/10.1016/j.eja.2024.127142

9. Bhuvaneswari S, Sharma SK, Punitha P, Shashidhar KS, Naveenkumar KL, Prakash N. Evaluation of morphological diversity of field pea [*Pisum sativum* subsp. *arvense* (L.)] germplasm under sub-tropical climate of Manipur. Legume Res. 2017;40(2):215-23. https://doi.org/10.18805/lr.v0iOF.10756

10. Srivastava A, Sharma A, Kumar R. Cluster analysis in field pea (*Pisum sativum* L.). J Pharmacogn Phytochem. 2018;7:655-7. https://doi.org/10.20546/ijcmas.2018.702.069

11. Griffing B. Concept of general and specific combining ability in relation to diallel crossing systems. Aust J Biol Sci. 1956;9(4):463-93. https://doi.org/10.1071/BI9560463

12. Azam MG, Sarker U, Hossain MA, Mahabubul Alam AKM, Islam MR, Hossain N, et al. Phenotypic diversity in qualitative and quantitative traits for selection of high yield potential field pea genotypes. Sci Rep. 2024;14(1):18561. https://doi.org/10.1038/s41598-024-69448-7

13. Prasad D, Nath S, Lal K, Jaiswal A, Yadav VP. Evaluation of field pea (*Pisum sativum* L.) genotypes for genetic variability and divergence. Int J Chem Stud. 2018;6(6):2288-91.

14. Ertiro TA. Genetic diversity based on cluster and principal component analyses for quantitative traits in field pea (*Pisum sativum* L.) genotypes at Arsi Highlands of Ethiopia. 2021. https://doi.org/10.9734/ajrcs/2022/v7i4142

15. Sanwal SK, Singh B, Singh V, Mann A. Multivariate analysis and its implication in breeding of desired plant type in garden pea (*Pisum sativum*). Indian J Agric Sci. 2015;85(10):1298-302. https://doi.org/10.56093/ijas.v85i10.52263

16. Kosev VI. Multivariate analysis of spring field pea genotypes. Banat J Biotechnol. 2015;6(11). https://doi.org/10.7904/2068-4738-VI(11)-23

17. Khan MRA, Mahbub MM, Reza MA, Shirazy BJ, Mahmud F. Selection of field pea (*Pisum sativum* L.) genotypes through multivariate analysis. Sci Agric. 2016;16(3):98-103. https://doi.org/10.15192/PSCP.SA.2016.16.3.98103

18. Habtamu S, Million F. Multivariate analysis of some Ethiopian field pea (*Pisum sativum* L.) genotypes. Int J Genet Mol Biol. 2013;5(6):78-87. https://doi.org/10.5897/IJGMB2013.0080

19. Hanci F, Cebeci E. Determination of morphological variability of different *Pisum* genotypes using principal component analysis. Legume Res. 2018;42(2):162-7. https://doi.org/10.18805/LR-438

20. Singh SR, Ahmed N, Singh DB, Srivastva KK, Singh RK, Mir A. Genetic variability determination in garden pea (*Pisum sativum* L. subsp. *hortense* Asch. and Graebn.) by using the multivariate analysis. Legume Res. 2017;40(3):416-22. https://doi.org/10.18805/lr.v0i0.7300