**Genetic Divergence of Principle Component Analysis and Cluster Analysis in Different Genotype Lentil (*Len culinaris L*.)**

**Abstract**
**Aims:**
The present study aimed to assess the genetic diversity among 25 lentil (Lens culinaris L.) genotypes, including three standard checks, with a focus on identifying high-yielding lines and those with elevated iron and zinc content, to support future breeding programs.

**Study Design:** The experiment was laid out in a randomized block design (RBD) with three replications.

**Place and Duration of Study:** The study was conducted at the Research Farm of Banda University of Agriculture and Technology, Banda, Uttar Pradesh, India, during the rabi season of 2023–2024.

**Methodology:** Twenty-five lentil genotypes, along with three checks (PL-04, PSL-9, and IPL-220), were evaluated for various agronomic and nutritional traits. Analysis of variance (ANOVA) was performed to determine genotypic variation. Cluster analysis and principal component analysis (PCA) were used to study genetic diversity and relationships among the genotypes.

**Results:** Significant genotypic differences were observed for most traits, indicating ample genetic variability. Cluster analysis grouped genotypes into three distinct clusters, with Cluster II containing the most genotypes (15), followed by Clusters I and III (5 each). The highest inter-cluster distance was found between Clusters II and III, suggesting the potential for generating superior recombinants through inter-cluster hybridization. PCA revealed four principal components accounting for 72.30% of total variance. The highest seed yield was recorded in genotypes ILL-753/ILL-8461 (0.294 kg/plot), ILL-7537/ILL-800-S4 (0.280 kg/plot), and Black lentil (0.284 kg/plot). Genotypes IPL-316 (119.5 mg/kg) and PL-4 (106.3 mg/kg) had the highest iron content, while X20115-89-23-S4 (54.4 mg/kg) and ILL-10657 (52.1 mg/kg) had the highest zinc content.

**Conclusion:** The observed genetic diversity among the lentil genotypes offers valuable resources for breeding programs. Genotypes with high yield and elevated micronutrient content can serve as promising parents for developing improved, high-yielding, and early-maturing lentil cultivars.

**Keywords:** Analysis of variance, Correlation, Cluster analysis

1. **Introduction**

Lentil (*Lens culinaris L. Medikus*), a member of the Leguminosae family is a diploid (2n = 2x = 14), self-pollinated, and annual cool season grain legume with a genome size of 4,063 Mpb is an important food component in most parts of the world (Faris *et al*., 2013). It is among the earliest domesticated plant species from the Middle East (Cokkizgin & Shtaya, 2013), having been cultivated for 10,000 years. The total world production of lentil is 5.610 million tons, the harvested area is 5.585 million hac, India’s total production of 1.490 million tons, harvested area is 1.734 million ha (FAOSTAT, 2023). Lentil has become an important rotational crop for wheat (Triticum aestivum L.) production systems in dryland regions. The inclusion of lentil makes grass weed control easier, breaks disease cycles, improves water use efficiency, and enhances yield (Chen et al., 2012; Miller et al., 2015), as well as offers the opportunity for improved economic returns (Tanaka et al., 2005). Iron (Fe) is an important part of hemoglobin, a protein that carries oxygen to body tissues. Globally, an estimated 50% of anemia cases are likely attributed to Fe deficiency (Stoltzfus, 2003), which makes the body more susceptible to other diseases (Hassan et al., 2018). About 33% of the world’s population suffers from zinc (Zn) deficiency, characterized by growth retardation and impaired immune function (Prasad, 2004). Lentil grains grown at multiple environments varied in total Fe and Zn concentrations among lentil genotypes, and with estimated Fe and Zn, respectively (Thavarajah et al., 2009). Significant cultivars and variations were observed in analyzing the micronutrient Fe and Zn content of lentil Genotype (Ray et al., 2014). A lentil genotype indicated significant heterogeneity in the grain Fe and Zn contents. Understanding the genetic basis of developmental and yield-related traits is vital for improving crop performance through breeding. This involves estimating key genetic parameters such as variance components, genotypic and phenotypic coefficients of variation, heritability, and genetic gain. Since small differences in quantitative traits can be difficult to distinguish visually, breeders often focus on traits that show clearer variation and are easier to select for. Correlation analysis is commonly used to evaluate the relationships among traits, providing insights into how different characteristics are linked to productivity. When traits associated with yield are well-characterized, they can serve as reliable selection criteria (Poehlman, 1991; Singh et al., 1995). Research has shown that genotypic correlations often exceed their phenotypic counterparts, suggesting stronger genetic linkages that may be masked by environmental factors. For example, considerable genetic and phenotypic variability in seed yield per plant. In contrast, Bicer and Sarkar (2004) found low heritability for several traits, including flowering time, seed yield, biomass, maturity, seed weight, pod number, and plant height. Conversely, Gowda et al. (1997) documented higher heritability and genetic advance for seed yield, indicating a stronger potential for genetic improvement. Traits such as hundred-seed weight, harvest index, and plant height have also shown positive and significant associations with seed yield in lentil. To optimize breeding efforts, it is essential to thoroughly evaluate the genetic diversity within lentil populations. This enables the effective selection of parent lines, facilitates the introduction of beneficial traits into elite varieties, and supports the long-term conservation of valuable genetic resources. Selecting genetically diverse parents for hybridization is particularly important, as such crosses increase the likelihood of producing superior offspring in future generations.

1. **Materials and Methods**

The experiment was conducted to evaluate 25 lentil genotypes, including three check varieties, under irrigated conditions and normal soil using a Randomized Block Design (RBD). Field trials were carried out during the Rabi season of 2023–2024 at the Postgraduate Research Block of the College of Agriculture, Banda University of Agriculture and Technology, Banda, India. The experimental site is located at an elevation of 113 meters above mean sea level, between latitudes 24°53′ and 25°55′ North and longitudes 80°07′ and 81°34′ East. Each genotype was sown in a plot consisting of three rows, with each row measuring 2 meters in length and spaced 30 cm apart, covering a total plot area of 1.2 m². Data were recorded from five randomly selected plants per genotype for the traits under study. Statistical analyses, including analysis of variance (ANOVA), cluster analysis, and principal component analysis (PCA), were performed to assess genetic variability and relationships among the genotypes.

1. **Results and Discussion**

**3.1. Analysis of variance**

Genetic diversity and effective selection approaches are fundamental to the success of any breeding program, as they enable the utilization of available genetic variability (23, 24). The analysis of variance (ANOVA) revealed statistically significant differences among the pigeon pea genotypes for nearly all the traits examined (Table 1), suggesting the presence of substantial inherent genetic variability within the population. This variability offers ample opportunity to identify genotypes with favorable traits for yield improvement, particularly when appropriate selection pressure is applied. These findings are consistent with previous reports by (Alemayo et al 2021 and Kumar et al. 2022). Significant genotypic differences were observed for key agronomic and nutritional traits, including days to 50% flowering (DF), number of primary branches (PB), number of secondary branches (SB), plant height (PH), number of pods per plant (PP), seeds per pod (SP), days to 50% maturity (DM), biological yield (BY), 1000-grain weight (TGW), iron content (Fe, mg/kg), zinc content (Zn, mg/kg), and seed yield (SY).

**Table 1. An analysis of variance (ANOVA) was conducted for twelve morphological and yield-related traits across lentil genotypes.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Source**  | **Replication**  | **Genotype** | **Error** |
| **df** | 2 | 24 | 48 |
| **Days to 50% flowering**  | 10.253 | 23.27\* | 10.587 |
| **Primary branches per plant** | 0.2125 | 0.5476\*\* | 0.09677 |
| **Secondary branches per plant** | 3.5897 | 7.0756\*\*\* | 0.8514 |
| **Plant height (cm)** | 49.204 | 38.44\* | 9.217 |
| **Total no. of pods per plant**  | 1089.5 | 29950.8\*\*\* | 333.1 |
| **Total no. of seed per pod**  | 0.011111 | 0.042133\* | 0.013244 |
| **Days to maturity**  | 18.8333 | 0.2133 | 1.13 |
| **Biological yield (kg)** | 0.03824 | 0.68931\*\*\* | 0.01451 |
| **1000 seed wt. (gm)**  | 103.455 | 43.944\*\* | 7.312 |
| **Fe** | 774.52 | 262.25\* | 71.05 |
| **Zn** | 93.545 | 33.739\*\*\* | 0.61 |
| **Seed yield (gm/plot)** | 0.003072 | 0.150492\*\*\* | 0.001786 |

*\*Indicates significance at 5%; Days to 50% flowering (DF), Number of primary branches (PB), Number of secondary branches (SB), Plant height (PH), Number of pods per plant (PP), Seeds per pod (SP), Days to 50% maturity (DM), Biological yield (BY), 1000-grain weight (TGW), Iron content (Fe, mg/kg), Zinc content (Zn, mg/kg), and Seed yield (SY).*

**3.2. Distribution of genotypes intoclusters**

To evaluate the genetic divergence among 25 genotypes genotypes, Mahalanobis' D² statistics were calculated for all possible genotype pairs. This analysis facilitated the assessment of genetic diversity within the population. Based on the results of cluster analysis, the genotypes were classified into three distinct and non-overlapping clusters. Among these, Cluster II contained the highest number of fifteen genotypes, followed by Clusters I and III, each comprising five genotypes (Table 2). In a separate analysis, 25 lentil genotypes were grouped into three clusters using the dendrogram method (Figure 1). Genotypes within the same cluster exhibited greater genetic similarity, whereas those from different clusters showed broader genetic divergence. Therefore, selecting parents from separate clusters could be advantageous for breeding programs, as such crosses have a higher potential to produce genetically diverse and heterotic offspring.

**Table 2. The clustering pattern of twenty-five genotypes, including three check varieties.**

|  |  |  |
| --- | --- | --- |
| **Cluster** | **Genotype number** | **Genotype Name** |
| **I** | 5 | X20125-109-513, 6002/LiRL-21-50-1-1/17-S3-S1, 590/8461/2-4-S1, ILL-7537/ILL-800-S4, X20115-89-23-S4 |
| **II** | 15 | IPL-406, IPL-316, PDL-1, ILL-753/ILL-8461, L-4076, ILL-6778XILL-5480-S3-S4, 4605/4380/2SPS, L-4147, ILL-10657, PL-04, L-4729, PSL-9, KM-1, IPL-220, Black lentil |
| **III** | 5 | 10072/1712/4-1, 6002/LiRL-21-50-1-1-1/24-6, X20125-146-S1, X20125-171-17-S1, 6002/7716/4-4-S2 |

* 1. **Inter-cluster and Intra-cluster distance**

The intra and inter-cluster distances between all possible pairs of three clusters were computed and presented in Table 3. While the minimum inter-cluster distance demonstrated a close link between the groupings, the largest inter-cluster distance indicated broad diversity. The maximum intra-cluster distance was found in Cluster III (4.580) followed by Cluster II (4.227), and the minimum intra-cluster distance was found in Cluster I (3.793). The maximum inter-cluster distance was found between cluster III (5.445), and the minimum inter-cluster distance was found between II (4.990). Similar results of utilization of principal component analysis combined with hierarchical cluster analysis in genetic diversity studies were reported by Rao, (Maurya *et al.* 2018) in lentil.

**Table 3. The average intra and inter-cluster distances for f clusters**

|  |  |  |  |
| --- | --- | --- | --- |
| **Cluster** | I | II | III |
| I | **3.793105** | 4.990322 | 5.088511 |
| II |   | **4.227174** | 5.445513 |
| III |   |   | **4.580984** |

*\*Bold values in the table represent intra-cluster distances*.

* 1. **Cluster Means for Different Characters**

The cluster means for all twelve traits are shown in Table 4. Among the clusters, Cluster II exhibited the highest mean values for several traits, including the number of pods per plant (120.77), primary branches per plant (3.79), and seeds per pod (1.93). In contrast, Cluster I had the highest mean values for iron content (94.60) and zinc (53.83). Cluster III displayed maximum values for biological yield (0.76), thousand seed weight (86.76), and overall yield (0.21), while Cluster I had the lowest values for these traits in Table 4. The results indicate that no single cluster contained genotypes with all desirable traits. Therefore, crossing genotypes from different clusters is recommended to develop superior genotypes.

**Table 4. Cluster Means for Different Characters in Lentil Germplasm.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Cluster** | **I** | **II** | **III** |
| **DF** | 55.46\* | 58.50 | 61.00\*\* |
| **PB** | 3.51\* | 3.79\*\* | 3.50 |
| **SB** | 7.21 | 8.41\*\* | 6.76\* |
| **PH** | 36.98\* | 42.99\*\* | 42.64 |
| **PP** | 78.74\* | 120.77\*\* | 116.09 |
| **SP** | 1.87\* | 1.93\*\* | 1.91 |
| **DM** | 105.29\* | 107.53\*\* | 110.24 |
| **BY** | 0.66\* | 0.69 | 0.78\*\* |
| **TQW** | 25.47 | 23.86\* | 29.79\*\* |
| **Fe** | 94.60\*\* | 87.02 | 86.76\* |
| **Zn** | 53.83\*\* | 46.07\* | 51.47 |
| **SY** | 0.21\* | 0.22 | 0.26\*\* |

*\*Lowest value, and \*\*Highest value*



**Figure 1. Dendrogram showing the relationship between seventy-three genotypes including three checks of pigeon pea genotypes.**

* 1. **Principal component analysis**

This method was first developed by Pearson (1901) and later utilized by Hotelling (1936). Several researchers are currently using it to select superior genotypes (Vianna *et al.*  2013 and Li *et al.* 2020). Principal Component Analysis (PCA) is a powerful tool in modern data analysis, as it is a well-known multivariate statistical technique used to identify the minimum number of components that can rank genotypes based on PC scores and explain the most variability. Principal components are generally estimated from either a correlation matrix or a covariance matrix. Considering the importance of PCA, this study focuses on lentil genotypes to identify the yield and yield-related traits responsible for the differences in yield among these genotypes. The estimates from the principal component analysis for the quantitative traits are presented in Table 5. The table reveals that the data from twelve traits were transformed into twelve principal components. The first four principal components, with eigenvalues greater than one, accounted for 72.30% of the total variation present in the studied genotypes. The first principal component accounted for the highest variation (25.43%), followed by PC2 (43.08%) and PC3 (58.76%). The eigenvalues for the first four principal components are as follows: the first (3.05), the second (2.12), the third (1.88), and the fourth (1.63). The traits contributing to the variance for each principal component are shown in Table 5.

**Table 5. Principal Component Analysis for Twelve Quantitative Traits in Lentil Genotypes**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Charcter** | **PC I** | **PC II** | **PC III** | **PC IV** |
| **Eigenvalue**  | 3.05 | 2.12 | 1.88 | 1.63 |
| **Variance (%)**  | 25.43 | 17.65 | 15.68 | 13.55 |
| **Cumulative variance (%)** | 25.43 | 43.08 | 58.76 | 72.31 |
| **Eigenvectors** |
| **DF** | **0.409** | -0.333 | 0.241 | 0.185 |
| **PB** | 0.345 | **0.438** | 0.021 | 0.151 |
| **SB** | 0.259 | **0.514** | 0.101 | 0.107 |
| **PH** | **0.464** | 0.044 | 0.003 | 0.172 |
| **PP** | **0.305** | -0.117 | 0.259 | -0.475 |
| **SP** | **0.172** | -0.090 | -0.119 | -0.565 |
| **DM** | **0.321** | -0.362 | 0.263 | 0.222 |
| **BY** | **0.178** | -0.257 | -0.561 | -0.051 |
| **TQW** | 0.012 | -0.335 | 0.174 | **0.239** |
| **Fe** | -0.058 | 0.030 | -0.262 | **0.481** |
| **Zn** | -0.370 | -0.238 | **0.249** | 0.114 |
| **SY** | **0.183** | -0.208 | -0.552 | 0.062 |

Principal Component I (PC I), which produced the most variability, was associated with days to 50% flowering, plant height, pods per plant, seeds per pod, days to maturity, biological yield, and seed yield per plot. The number of primary branches and secondary branches mainly contributed to PC II, while zinc content was significant in PC III. In PC IV, 1000 seed weight and iron content were prominent (Table 5). These results indicate that these traits contributed most to the divergence and accounted for the majority of the variability. Thus, the yield level was determined by these four components, representing a weighted average of the traits. The decision on how many components to retain is often based on the eigenvalues. Typically, the number of variables corresponds to the sum of the eigenvalues. In this investigation, only the first four principal components (PCs) exhibited eigenvalues greater than 1.00 and demonstrated maximum variability among lentil germplasm regarding yield component traits. The traits associated with these four PCs should be prioritized in lentil improvement programs.

**Figure 2. PCA Biplot of 25 genotypes including three checks of lentil genotypes for yield and yield-related traits.**

Principal Component Analysis (PCA) is an effective multivariate technique for identifying and determining the independent principal components that govern plant attributes individually. It can be used to select effective traits through indirect selection of superior genotypes. The gaps between traits indicate their associations; for instance, if two variables are positioned away from the origin and form an acute angle (less than 90°), they are positively correlated. This is exemplified by biological yield, as indicated by PC1 and PC2, which reflects the contribution of these traits toward divergence (Figure. 2). The length of the vector represents the magnitude of each character. The scree plot shows that the major contributions to divergence were primarily from PC1, followed by PC2, PC3, and PC4. A scatter plot of twelve characters using PC1 and PC2 is presented in (Fig. 2). The Scree plot reflected the eigenvalue of different principal components and the percent cumulative variability as shown by PCA of lentil genotype based on quantitative traits in Figure 3.

**Figure 3. The scree plot shows the eigenvalues and cumulative variability of principal components in the PCA of lentil genotype.**

**Figure 4. Scatter plot of thirteen quantitative characters of lentil using PC1 and PC2.**

**Conclusion**

The present study revealed significant genetic diversity among 25 lentil genotypes and three checks, indicating a rich source of variability that can be effectively harnessed in lentil improvement programs. The significant genotypic variation observed for most traits suggests the presence of ample genetic potential for enhancing seed yield and micronutrient content, particularly iron and zinc, through selective breeding. Cluster analysis divided the genotypes into three distinct groups, with the greatest inter-cluster distance observed between Clusters II and III, indicating the possibility of producing high-quality recombinants through inter-cluster hybridization. Principal component analysis further supported the diversity, with four components explaining over 72% of the total variation. High-yielding genotypes such as ILL-753/ILL-8461, ILL-7537/ILL-800-S4, and Black lentil lines, as well as nutrient-rich genotypes like IPL-316 and X20115-89-23-S4, demonstrated their potential as valuable genetic resources. These findings provide a strong foundation for selecting elite parental lines to develop improved cultivars with superior yield potential and enhanced nutritional quality. The integration of such diverse and high-performing genotypes into breeding programs could contribute significantly to food and nutritional security, particularly in regions dependent on lentils as a primary protein and micronutrient source. Future efforts should focus on utilizing these lines in strategic hybridization and selection schemes.

**Compliance with ethical standards**

**Ethical issues:** None

**References**

Alemayo, G. T., Gurmu, G. N., Singh, B. C. S., & Getachew, Z. E. (2021). Genetic variability of chickpea (*Cicer arietinum* L.) genotypes for yield and yield components in West Shewa, Ethiopia. *Plant Cell Biotechnology and Molecular Biology, 22*(35–36), 307–331.

Bicer, B. T., & Sarkar, D. (2004). Evaluation of some lentil genotypes at different locations in Turkey. *International Journal of Agriculture and Biology, 6*(2), 317–332.

Chen, C., Etemadi, F., Franck, W., Franck, S., Abdelhamid, M. T., Ahmadi, J., Mohammed, Y. A., Lamb, P., Miller, J., Carr, P. M., & McPhee, K. (2022). Evaluation of environment and cultivar impact on lentil protein, starch, mineral nutrients, and yield. *Crop Science, 62*(2), 893–905.

Cokkizgin, A., & Shtaya, M. J. (2013). Lentil: Origin, cultivation techniques, utilization and advances in transformation. *Agricultural Science, 1*(1), 55–62.

Food and Agriculture Organization. (2023). *World food and agriculture – Statistical yearbook 2023*. Rome. <https://doi.org/10.4060/cc8166en>

Faris, M. E. A. I. E., Takruri, H. R., & Issa, A. Y. (2013). Role of lentils (*Lens culinaris* L.) in human health and nutrition: A review. *Mediterranean Journal of Nutrition and Metabolism, 6*(1), 3–16.

Gowda, M. B., Prakash, J. C., & Hala, J. S. (1997). Estimation of genetic variability and heritability in black gram (*Vigna mungo* L. Hepper). *Crop Research, 13*(2), 369–372.

Hassan, S., Sofi, P. A., Khuroo, N. S., Rehman, K., & Bhat, M. A. (2018). Natural variation for root traits, morphological and yield parameters in lentil (*Lens culinaris* M.). *Bulletin of Environment, Pharmacology and Life Sciences, 7*, 80–84.

Kumar, K., Anjoy, P., Sahu, S., Durgesh, K., Das, A., Tribhuvan, K. U., et al. (2022). Single trait versus principal component-based association analysis for flowering related traits in pigeon pea. *Scientific Reports, 12*(1), 10453.

Li, X., Zhang, C., Behrens, H., & Holtz, F. (2020). Calculating amphibole formula from electron microprobe analysis data using a machine learning method based on principal components regression. *Lithos, 362*, 105469.

Maurya, K. S., Kumar, S., Choudhary, A. M., & Purushottam, V. S. M. (2018). Diversity analysis in lentil (*Lens culinaris* Medik.). *Journal of Pharmacognosy and Phytochemistry, 7*, 1958–1961.

Miller, P. R., Atencio, S. C., Jones, C. A., Carr, P. M., Eriksmoen, E., Franck, W., Rickertsen, J., Fordyce, S. I., Ostlie, M., Lamb, P. F., & Fonseka, D. L. (2024). Inoculant and fertilizer effects on lentil in the US northern Great Plains. *Agronomy Journal*.

Poehlman, J. M. (1991). *The mungbean*. Oxford and IBH Publishing Co. Pvt. Ltd.

Prasad, A. S., Bao, B., Beck, F. W., Kucuk, O., & Sarkar, F. H. (2004). Antioxidant effect of zinc in humans. *Free Radical Biology and Medicine, 37*(8), 1182–1190.

Ray, H., Bett, K., Tar’an, B., Vandenberg, A., Thavarajah, D., & Warkentin, T. (2014). Mineral micronutrient content of cultivars of field pea, chickpea, common bean, and lentil grown in Saskatchewan, Canada. *Crop Science, 54*(4), 1698–1708.

Singh, K. P., & Singh, V. P. (1995). Comparative role of seed yield components in mungbean (*Vigna radiata* L. Wilczek). *Legume Research, 18*, 109–112.

Stoltzfus, R. J. (2003). Iron deficiency: Global prevalence and consequences. *Food and Nutrition Bulletin, 24*(4 Suppl 2), S99–S103.

Tanaka, D. L., Krupinsky, J. M., Merrill, S. D., Liebig, M. A., & Hanson, J. D. (2007). Dynamic cropping systems for sustainable crop production in the northern Great Plains. *Agronomy Journal, 99*(4), 904–911.

Thavarajah, D., Thavarajah, P., Sarker, A., & Vandenberg, A. (2009). Lentils (*Lens culinaris* Medikus Subspecies *culinaris*): A whole food for increased iron and zinc intake. *Journal of Agricultural and Food Chemistry, 57*(12), 5413–5419.

Vianna, V. F., Unêda-Trevisoli, S. H., Desidério, J. A., Santiago, S. D., Charnai, K., & Ferreira Júnior, J. A. (2013). The multivariate approach and influence of characters in selecting superior soybean genotypes. *African Journal of Agricultural Research, 8*(30), 4162–4169.