



Multivariate Dissection of Yield-associated Traits in Greengram (*Vigna radiata* L. Wilczek) Using Principal Component, Path and Cluster Analysis

Venkata Raja V. Mokshith¹, V. Roja², N. H. Satyanarayana³, K. Jayalalitha⁴ and P. Supriya⁵

¹Dept. Molecular Biology and Biotechnology, S. V. Agricultural College, ANGRAU, Tirupati, Andhra Pradesh (517 502), India

²Regional Agricultural Research Station, ANGRAU, Maruteru, Andhra Pradesh (534 122), India

³Regional Agricultural Research Station, ANGRAU, Lam, Guntur, Andhra Pradesh (522 034), India

⁴Dept. of Plant Physiology, Agricultural College, ANGRAU, Bapthla, Andhra Pradesh (522 101), India

⁵ICM Division, ICAR-NAARM, Rajendranagar, Hyderabad, Telangana (500 030), India



Corresponding ✉ vmokshith.msc@gmail.com

 0009-0007-3522-1726

ABSTRACT

The experimental study was conducted using 83 greengram genotypes during the *rabi* season (November, 2023 to February, 2024) at the Regional Agricultural Research Station (RARS), Lam, Guntur, Andhra Pradesh, India to study the yield-related traits and employed an integrated analytical framework comprising path coefficient analysis, principal component analysis (PCA), and hierarchical cluster analysis with heatmap visualization. Path coefficient analysis revealed that pods plant⁻¹ and days to maturity exerted a direct effect on the yield trait, indicating that these traits played a crucial role in influencing seed yield. In total, seven principal components were extracted, of which the first three PCs exhibited greater than 1 and collectively accounted for approximately 82.37% of total variability, suggesting that a substantial proportion of genetic diversity was captured within these components. The 83 genotypes were classified into 4 groups through hierarchical clustering. Group III comprised the largest number of genotypes (35), followed by Group I (25 genotypes) and Group II (21 genotypes), while Group IV contained only 2 genotypes. Additionally, a two-way clustered heatmap further elaborated on genotype-trait interactions and facilitated the identification of high-performing subsets for specific attributes. Genotypes belonging to group III demonstrated superior performance for pods plant⁻¹ and seed yield plant⁻¹ (g), while group I genotypes excelled in early maturity traits, making both groups promising candidates for targeted selection in future greengram improvement programmes.

KEYWORDS: Greengram, path coefficient, principal component analysis, hierarchical clustering

Citation (VANCOUVER): Mokshith et al., Multivariate Dissection of Yield-associated Traits in Greengram (*Vigna radiata* L. Wilczek) Using Principal Component, Path, and Cluster Analysis. *International Journal of Bio-resource and Stress Management*, 2026; 17(8), 01-06. [HTTPS://DOI.ORG/10.23910/1.2026.7122](https://doi.org/10.23910/1.2026.7122).

Copyright: © 2026 Mokshith et al. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, that permits unrestricted use, distribution and reproduction in any medium after the author(s) and source are credited.

Data Availability Statement: Legal restrictions are imposed on the public sharing of raw data. However, author have full right to transfer or share the data in raw form upon request subject to either meeting the conditions of the original consents and the original research study. Further, access of data needs to meet whether the user complies with the ethical and legal obligations as data controllers to allow for secondary use of the data outside of the original study.

Conflict of interests: The author has declared that no conflict of interest exists.

RECEIVED on 13th April 2026

RECEIVED in revised form on 10th July 2026

ACCEPTED in final form on 26th July 2026

PUBLISHED on 05th August 2026

1. INTRODUCTION

Greengram (*Vigna radiata* (L.) Wilczek.) which is also known as poor man's meat (Sudhakaran and Bukkan, 2021, Goud et al., 2022 and Kumar et al., 2023), is a short-duration, self-pollinated, diploid legume species widely cultivated for its high protein content, fast growth, and adaptability to wide agroclimatic conditions (Pathak et al., 2023). Along with high protein content, it is rich in complex carbohydrates, dietary fiber, minerals such as iron and zinc, and various bioactive compounds (Poutanen et al., 2022). It has a keen role in soil fertility restoration through fixation of atmospheric nitrogen (N_2) through symbiotic association with *Rhizobium* species in the soil (Wanga et al., 2017, and Kaur et al., 2021). Being a warm-season crop with a short period of 60–90 days, greengram fits well into diverse cropping systems and serves as an ideal intercrop and rotation crop, thereby contributing significantly to sustainable agricultural practices across tropical and sub-tropical regions. Even though greengram provides various nutritional and agricultural benefits, its yield has not significantly improved due to many factors like limited genetic diversity, vulnerability to various biotic and abiotic stresses, and limited adoption of improved varieties (Asghar et al., 2021). Nearly 11% of the total pulse production and 16% of the total pulse area in India is contributed by greengram (Mokshith et al., 2025). Genetic improvement of the greengram focuses on generating genetic material, selecting superior genotypes from diverse genetic sources, and developing improved varieties, which involves agronomical and morphological characterization of germplasm. Even though advancements were made in greengram breeding, challenges like narrow genetic diversity and yield instability remained unsolved. Although several varieties of greengram were developed, most share a similar genetic makeup. To create high-yielding, superior cultivars with various desirable traits, it is essential to incorporate highly diverse genotypes in the breeding program. Utilizing modern tools like Principal Component Analysis (PCA), cluster analysis, and path analysis, along with traditional methods, will be required to address these issues. Path analysis gives us a detailed understanding of how different traits directly and indirectly contributes to yield. Principal component analysis (PCA) is a multivariate statistical tool to identify the minimum number of components that can explain the maximum variability out of the total variability and also rank genotypes based on PC scores (Anderson, 1972). PCA is used in germplasm characterization features, to visualize individual differences and relationships, and assess their contribution to total variation (Singh et al., 2016). This approach is beneficial when screening a large number of genetic resources with a large number of descriptor variables (Beiragi et al., 2011, and Golbashy et al., 2010).

Earlier studies by Tahir et al. (2020), Khatun et al. (2022), Nalajala et al. (2023) were based on PCA for studying a large number of germplasms. Clustering analysis further helps in identifying genetically diverse genotypes that can be used as parent lines in crop improvement programs and enables us to exploit trait combinations for specific agro-ecological zones for enhancing adaptability and yield stability. Studies of Asadullah et al. (2024), Sridhar et al (2022a, b), and Das et al. (2026), etc., successfully grouped genotypes into clusters based on their trait profiles, identifying superior genotypes for yield and yield-related traits. Keeping these points in view, the present study was undertaken to evaluate the 83 genotypes for yield and yield-related traits through these multivariate approaches.

2. MATERIALS AND METHODS

This experimental study was carried out using 83 greengram genotypes during the rabi season (November, 2023 to February, 2024) at the Regional Agricultural Research Station (RARS), Lam, Guntur (31.5 m above mean sea level, 16.18 N and 80.29 E), Andhra Pradesh, India. 83 diverse greengram genotypes were obtained from the AICRP on *kharif* Pulses center, RARS, Lam. These genotypes represent different agro-climatic zones of India. Each genotype was sown in 2 rows, each row was 4 m in length with a spacing of 30×10 cm² in an augmented randomized block design. Plant protection measures and intercultural operations were done at regular intervals during crop growth.

The characters that were studied viz., days to 50% flowering (number of days from sowing till 50% of plants in a genotype had atleast one open flower, days to maturity (number of days taken for 80% pods maturation on a plant from date of sowing), clusters plant⁻¹ (number of pod bearing clusters in a plant), pods plant⁻¹, seeds pod⁻¹, seed yield plant⁻¹ (g), and test weight i.e. 100 seed weight (g) were recorded on five randomly selected plants from each genotype. Statistical analysis was performed based on the mean values of the five plants' data. Principal component analysis (PCA), cluster analysis, and path analysis were all conducted using R software (Anonymous, 2025). “Facto MineR” and “ggfortify” for Principal Component Analysis (PCA), “cluster” and “dend” extend for cluster analysis, and “lavaan” and “semPlot” for path analysis were the R packages used.

3. RESULTS AND DISCUSSION

3.1. Path analysis

Path coefficient analysis revealed that days to maturity (0.961) and pods plant⁻¹ (0.561) exerted strong positive direct effects on seed yield. Clusters plant⁻¹ (0.269), seeds pod⁻¹ (0.180), and test weight (0.156) showed moderate to low positive direct contributions (Table 1). In contrast,

Table 1: Path analysis

Trait	DF	DM	CP	PP	SP	TW	direct Effect	Total effect (r)
DF	-	0.951	0.011	-0.030	0.003	-0.027	-0.959	-0.050
DM	-0.949	-	0.008	-0.029	0.002	-0.026	0.961	-0.032
CP	-0.040	0.027	-	0.400	0.029	-0.050	0.269	0.635
PP	0.051	-0.049	0.192	-	0.044	-0.033	0.561	0.765
SP	-0.017	0.013	0.043	0.136	-	-0.053	0.180	0.302
TW	0.168	-0.158	-0.087	-0.118	-0.061	-	0.156	-0.099

DF: Days to 50% flowering; DM: Days to maturity; CP: Clusters pod⁻¹; PP: Pods plant⁻¹; SP: Seed pod⁻¹; TW: Test weight

days to 50% flowering exhibited a high negative direct effect (-0.959). However, the total effect analysis indicated a different trend. Pods plant⁻¹ showed the highest positive association with seed yield (0.765), followed by clusters plant⁻¹ (0.635) and seeds pod⁻¹ (0.302). Days to maturity and days to 50% flowering showed negligible total effects due to the cancellation of their indirect influences. Test weight exhibited a slight negative total effect (-0.099). Among indirect effects, days to 50% flowering had a strong positive indirect influence via days to maturity (0.951), while clusters plant⁻¹ contributed substantially to yield through pods plant⁻¹ (0.400). Similarly, pods plant⁻¹ also exhibited positive indirect effects via clusters plant⁻¹ (0.192), indicating interdependence among yield components.

The residual effect (0.587) indicated that a moderate proportion of variability in seed yield was not explained by the traits included in the analysis. Traits test weight, clusters plant⁻¹, and seeds pod⁻¹ were shown to have a significant direct effect on seed yield (Dawane et al., 2022; Yadav et al., 2024). Overall, the results emphasized that yield improvement should focus primarily on pods plant⁻¹, followed by clusters plant⁻¹, while caution should be exercised when interpreting the effects of phenological traits due to their interdependence.

3.1. Principal component analysis (PCA) and clustering

Principal component analysis (PCA) was conducted to identify the major traits contributing to variability among the 83 greengram genotypes and to determine the interrelationships among the yield-related traits. Seven principal components were extracted in total, of which the first three PCs exhibited more than 1 eigenvalue and showed about 82.37% of total variability. The value of the eigenvalue and variations explained by each of the PC are depicted in Table 2.

The first principal component (PC1) showed 37.7% of the total phenotypic variance, representing the single largest source of diversity in the studied germplasm, followed by PC2 explaining 29.3%. Our present findings were consistent with earlier PCA studies of greengram, where

Table 2: Eigenvalues, proportion of variance, and cumulative variance of all principal component

PC	Eigen values	Proportion of variance	Cumulative proportion
PC1	1.6248	0.3771	0.3771
PC2	1.4330	0.2933	0.6705
PC3	1.0357	0.1532	0.8237
PC4	0.8514	0.1035	0.9273
PC5	0.5348	0.0409	0.9681
PC6	0.4624	0.0306	0.9987
PC7	0.0965	0.0013	1.0000

a small number of components capturing 70–85 % of total variation has been considered sufficient for describing the major phenotypic structure of genotypes (Nalajala et al., 2023; Jadhav et al., 2021). The scree plot (Figure 1) illustrated the percentage of explained variance across 7 principal components, exhibiting a gradual eigenvalue decline. These results were in agreement with the findings of Patel et al. (2021).

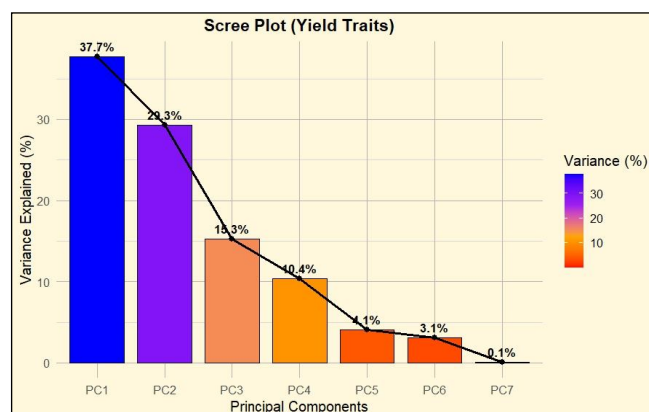


Figure 1: Scree plot of 7 principal components (PCs) against their eigenvalues

The PCA biplot (Figure 2) revealed substantial dispersion of genotypes and trait vectors onto the PC1-PC2 plane, providing an integrated view of genotype-trait relationships across the germplasm. Traits with long vectors exerted

greater influence on genotype differentiation, while the angular relationships among the vectors indicated the direction and strength of inter-trait associations. Genotypes positioned in close to specific trait vectors exhibited comparatively higher relative expression of those attributes, enabling visual identification of trait-superior genotypes. Genotypes situated at a distance from the central mass, i.e., periphery of the biplot, represent phenotypically distinct accessions with extreme trait profiles that might be useful for selecting as parents in crossing programs targeting specific trait combinations. To further resolve genotype relationships, hierarchical clustering was performed based on yield-related trait profiles using Euclidean distance as the dissimilarity measure and the unweighted pair-group method with arithmetic mean (UPGMA). The resulting dendrogram partitioned the genotypes into four distinct clusters at the selected dissimilarity threshold. Group III comprised the largest number of genotypes (35), followed by Group I (25) and Group II (21), while Group IV contained only 2 genotypes. Group III has the majority of genotypes, indicating broad phenotypic similarity within the population, reflecting the common genetic backgrounds or shared ancestry from similar agro-ecological zones of origin. The clustering topology was mostly consistent with PCA-based spatial distribution, reinforcing the robustness and consistency of the multivariate grouping.

A two-way clustered heatmap (Figure 3) constructed using independently normalized trait values (0–1 scaling) provided a high-resolution visualization of genotype-

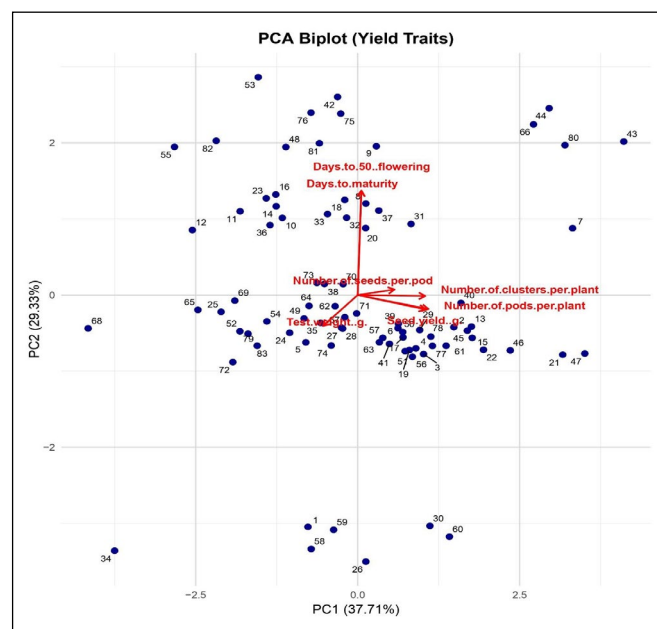


Figure 2: Biplot showing the distribution of genotypes and the contribution of all the traits in the first two principal components

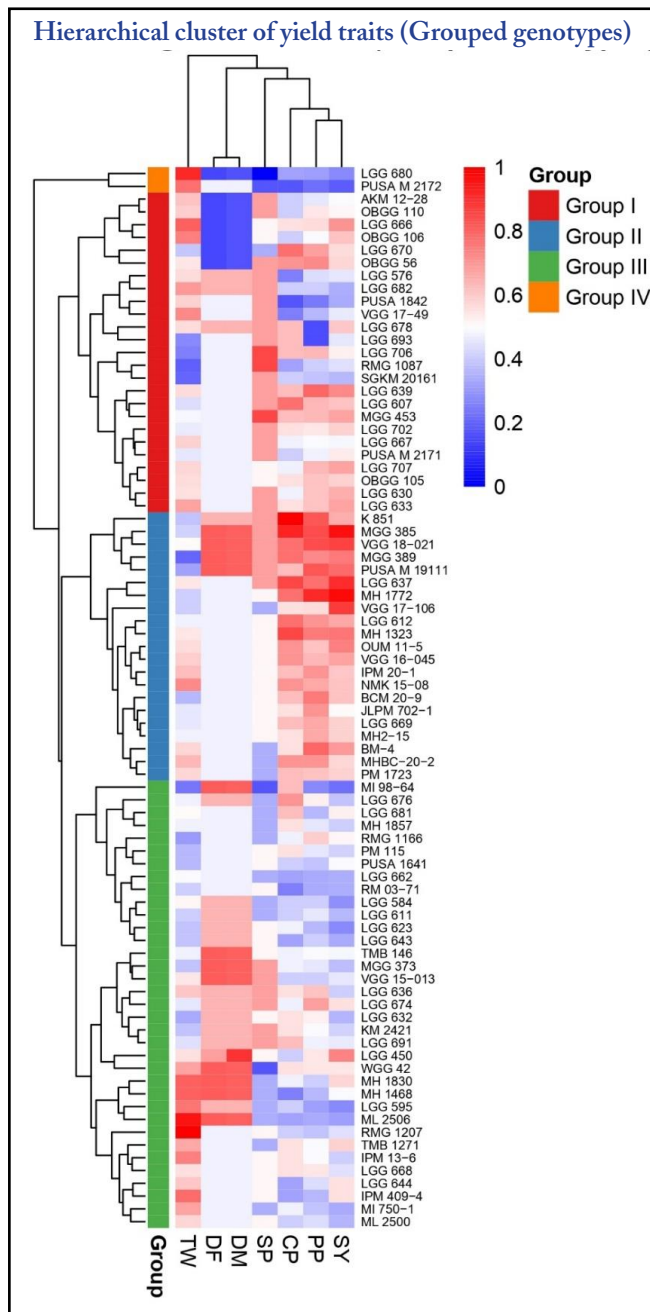


Figure 3: Clustered heatmap shows normalized trait values across 83 mungbean genotypes. For visualization, the values for each trait were independently scaled to a range of 0 (minimum value) to 1 (maximum value), where high values are shown in red and low values in blue. Both genotypes and traits were clustered using Euclidean distance and the Group Average (Unweighted Pair-Group) clustering method

trait interactions. The heatmap gradient from blue (low normalized value) to red (high normalized value) enabled the identification of high- and low-performing genotype subsets for each trait, enabling discrimination of elite and contrasting genotypes within the population.

4. CONCLUSION

Collectively, these multivariate analyses demonstrated substantial and well-structured phenotypic diversity among 83 greengram genotypes, underscoring the presence of distinct yield-performance architectures that could be strategically exploited in breeding programs.

5. REFERENCES

- Anderson, T.W., 1972. An introduction to multivariate analysis. Wiley Eastern Pvt. Ltd., New Delhi. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=1912798>.
- Anonymous, 2025. RStudio: Integrated development environment for R (Boston, MA: Posit Software, PBC). Available from: <http://www.posit.co/>. Accessed on: July 7, 2025.
- Asadullah, H.M., Malik, S.R., Khalid, R., Ajmal, S.A., Faisal, S., Abdullah, M., Ghuffar, S., 2024. Genetic analysis and breeding strategies for Mungbean (*Vigna radiata* L.): insights from correlation, principal component, and cluster analyses. *Planta Animalia* 3(1), 57–64. Available from: <https://plantanimalia.com/index.php/journal/article/view/32>.
- Asghar, M.J., Hameed, A., Shahid, M., Rizwan, M., Azeem, F., 2021. Mungbean wild genetic resource: a potential source of genetic improvement for biotic and abiotic stress tolerance. In: *Wild Germplasm for Genetic Improvement in Crop Plants*. Academic Press, 299–319. DOI: 10.1016/B978-0-12-822137-2.00016-3.
- Beiragi, M.A., Ebrahimi, M., Mostafavi, K., Golbashy, M., Khorasani, S.K., 2011. A study of morphological basis of corn (*Zea mays* L.) yield under drought stress condition using correlation and path coefficient analysis. *Journal of Cereals and Oilseeds* 2(2), 32–37. Available from: https://journalbackups.lon1.digitaloceanspaces.com/uploads/main/article/article1379170749_Beiragi%20et%20al.pdf.
- Das, S., Sawarkar, A., Saha, S., Raman, R.B., Dasgupta, T., 2026. Principal component and cluster analysis in mungbean [*Vigna radiata* (L.) Wilczek]. *Legume Research: An International Journal* 49(1). DOI: 10.18805/LR-5305.
- Dawane, B.K., Patil, D.K., Gite, V.K., Dawane, J.K., Kale, D.G., 2022. Character association and path analysis studies on seed yield and its yield attributing traits in greengram (*Vigna radiata* (L.) Wilczek). *The Pharma Innovation Journal* 11(12), 2476–2481. Available from: <https://www.thepharmajournal.com/archives/2022/vol11issue12/PartAE/11-12-46-501.pdf>.
- Golbashy, M., Ebrahimi, M., Khorasani, S.K., Choukan, R., 2010. Evaluation of drought tolerance of some corn (*Zea mays* L.) hybrids in Iran. *African Journal of Agricultural Research* 5(19), 2714–2719. Available from: https://academicjournals.org/article/article1380965816_Golbashy%20et%20al.pdf.
- Goud, V.V., Konde, N.M., Jaybhaye, C.P., Solunke, P.S., Kharche, V.K., 2022. Influence of potassium on growth, yield, water relation and chlorophyll content of green gram [*Vigna radiata* (L.) Wilczek] in inceptisols. *Legume Research* 45(12), 1523–1527. Available from: <https://www.arcjournals.com/journal/legume-research-an-international-journal/LR-4279>.
- Jadhav, R.A., Mehtre, S.P., Patil, D.K., Gite, V.K., 2021. Multivariate analysis using D2 and principal component analysis in Mung bean [*Vigna radiata* (L.) Wilczek] for study of genetic diversity. *Legume Research-An International Journal* 1(8). DOI: 10.18805/LR-4508.
- Kaur, L., Kaur, A., Brar, A.S., 2021. Water use efficiency of green gram (*Vigna radiata* L.) impacted by paddy straw mulch and irrigation regimes in north-western India. *Agricultural Water Management* 258, 107184. DOI: <https://doi.org/10.1016/j.agwat.2021.107184>.
- Khatun, R., Uddin, M.I., Uddin, M.M., Howlader, M.T.H., Haque, M.S., 2022. Analysis of qualitative and quantitative morphological traits related to yield in country bean (*Lablab purpureus* L. sweet) genotypes. *Heliyon* 8(12). DOI: 10.1016/j.heliyon.2022.e11631.
- Kumar, S., Mahapatra, M., Behera, D., Pradhan, P.L., Swain, S.K., Rath, I., Sahni, R.K., 2023. Effect of stem diameter, cutting speed and moisture content on cutting torque for green gram harvesting. *International Journal of Bio-resource and Stress Management* 14(8), 1127–1132. DOI: 10.23910/1.2023.3593b.
- Mokshith, V.R.V., Roja, V., Hari Satyanarayana, N., Jayalalitha, K., Supriya, P., 2025. Genetic variability and association analysis between yield and yield-related traits in greengram (*Vigna radiata* (L.) Wilczek). *International Journal of Bio-resource and Stress Management* 16(5), 01–08. DOI: <https://doi.org/10.23910/1.2025.6054>.
- Nalajala, S., Muniyandi, S.J., 2023. Principal component analysis of quantitative and qualitative traits in sixty mung bean (*Vigna radiata* L. Wilczek) genotypes. *International Journal of Bio-resource and Stress Management* 14(4), 643–651. Available from: <https://ojs.pphouse.org/index.php/IJBBSM/article/view/3722/1202>.
- Patel, K.V., Parmar, D.J., Kundaria, V.B., Patel, H.P., Patel, B.N., 2021. Tailoring genetic diversity of greengram genotypes through principal component and cluster analysis. *Electronic Journal of Plant Breeding* 12(1),

163. Available from: <https://openurl.ebsco.com/contentitem/gcd:149698837?sid=ebsco:plink:scholar&id=ebsco:gcd:149698837&crl=c>.
- Pathak, R., Panchariya, P., Choudhary, M., Solanki, K., Rani, R., Kakani, R.K., Kalia, R.K., 2023. Morphophysiological and molecular diversity in mung bean (*Vigna radiata* L.). In: Legumes: Physiology and Molecular Biology of Abiotic Stress Tolerance, 115–147. Available from: https://www.researchgate.net/publication/368227666_morphophysiological_and_molecular_diversity_in_mung_bean_vigna_radiata_l.
- Poutanen, K.S., Karlund, A.O., Gómez-Gallego, C., Johansson, D.P., Scheers, N.M., Marklinder, I.M., Landberg, R., 2022. Grains—a major source of sustainable protein for health. *Nutrition Reviews* 80(6), 1648–1663. DOI: 10.1093/nutrit/nuab084.
- Sridhar, V., Rao, S.S., Sriram, A., Gopal, G.V., 2022. Studies on genetic divergence among greengram (*Vigna radiata* L.) germplasm accessions. *International Journal of Bio-resource and Stress Management* 13(8), 838–844. Available from: <https://ojs.pphouse.org/index.php/IJBSM/article/view/4287>.
- Sridhar, V., Sriram, A., Rao, S.S., Gopal, G.V., 2022b. Evaluation of greengram (*Vigna radiata* L.) germplasm for seed yield and yield-related traits through cluster analysis. *International Journal of Environment and Climate Change* 12(11), 3563–3574. DOI: 10.9734/ijecc/2022/v12i111405.
- Singh, S., Prakash, A., Chakraborty, N.R., Wheeler, C., Agarwal, P.K., Ghosh, A., 2016. Trait selection by path and principal component analysis in *Jatropha curcas* for enhanced oil yield. *Industrial Crops and Products* 86, 173–179. DOI: <https://doi.org/10.1016/j.indcrop.2016.03.047>.
- Sudhakaran, S.M.N., Bukkan, D.S., 2021. A review on nutritional composition, antinutritional components and health benefits of green gram (*Vigna radiata* (L.) Wilczek). *Journal of Food Biochemistry* 45(6), e13743. DOI: <https://doi.org/10.1111/jfbc.13743>.
- Tahir, A., Ilyas, M.K., Sardar, M.M., Pouya, A.K., Rasouli, F., Bibi, A., Ghafoor, A., 2020. Selection criteria for yield potential in a large collection of *Vigna radiata* (L.) accessions. *Euphytica* 216(9), 138. DOI: 10.1007/s10681-020-02675-x.
- Wanga, L., Baib, P., Yuanc, X., Chena, H., Wanga, S., Chenc, X., Chenga, X., 2017. Genetic diversity assessment of a set of introduced mung bean accessions (*Vigna radiata* L.). *The Crop Journal* 6(2), 207–213. DOI: <https://doi.org/10.1016/j.cj.2017.08.004>.
- Yadav, K.K., Kumari, J., Mamatha, A., Kumar, V., Gantait, A., Maxton, A., 2024. Study of genetic variability and path analysis for yield in greengram (*Vigna radiata* L. Wilczek). *Journal of Experimental Agriculture International* 46(9), 130–141. DOI: 10.9734/jeai/2024/v46i92812.