Principal Component Analysis in Kodo Millet (*Paspalum scrobiculatum*) Under Salt Stress Conditions

ABSTRACT

The study was carried out in a polyhouse using germination trays to examine the impact of salt stress on seedling growth and germination across 42 Kodo Millet genotypes. Conducted at the Research Farm of RVSKVV, Gwalior, the experiment followed a Completely Randomized Design (CRD) with three replications. To minimize variability and maintain uniform conditions, a standardized growing medium composed of compost, vermiculite, and cocopeat in a 1:1:1 ratio was employed. Salt stress levels were assessed using NaCl solutions at concentrations of 50 mM, 100 mM, 250 mM, and 500 mM. In this study, two principal components with eigen values greater than one collectively accounted for 72.2% of the total variability among the analyzed traits. The first principal component (PC1) contributed the most, explaining 63.05% of the variation, while the second principal component (PC2) explained 9.15%. PC1 was primarily associated with traits such as days to emergence, final germination percentage, shoot length, seedling length, shoot fresh weight, root fresh weight, total fresh weight, total dry weight, and proline content. In contrast, PC2 was dominated by traits such as root length, vigor index, chlorophyll a, and chlorophyll b. The scores for PC1 ranged from 0.002 to 3.795 for positive values and -7.521 to -0.034 for negative values, while PC2 scores ranged from 0.061 to 2.894 for positive values and -2.687 to -0.089 for negative values. Scree plot and biplot) collectively highlighted the major contributors to variability and their interrelationships under the salt stress condition in kodo millet genotypes.

Keywords: kodo Millet, Principal Component Analysis, Salt Sress

1. INTRODUCTION

Kodo millet (*Paspalum scrobiculatum*) is increasingly important due to its resilience for biotic and abiotic stresses, making it ideal for regions facing climate change impacts [1] [2]. Its ability to grow with minimal inputs supports sustainable, low-input farming systems [3] [4]. Nutritionally, kodo millet is rich in protein, fiber, iron, and antioxidants, with a low glycemic index, making it beneficial for managing diabetes and promoting overall health [5] [6] [7] [8]. As a gluten-free grain, it also supports dietary diversity and addresses micronutrient deficiencies [9] [10]. Its role in ensuring food security, especially for smallholder farmers in marginal areas, and its contribution to biodiversity and sustainable agriculture make kodo millet a crucial crop in today's food systems.

Although kodo millet is generally tolerant to various stresses, its yield is adversely affected by salinity stress, as it is a glycophyte crop that can only tolerate low levels of salt [11] [12]. India incurs economic losses of approximately 249 billion Indian rupees each year due to around 6.73 million hectares of salt-affected soils (SAS) that remain un-reclaimed. These losses are projected to rise substantially, as future estimates suggest that SAS areas could expand to nearly 16 million hectares by 2050, driven by improper irrigation practices and the impacts of climate change [13]. Cultivating salt-tolerant crops and varieties is a key strategy for soil reclamation and improving

productivity [14]. Salt-tolerant cultivars are more effective in reclaiming salt-affected soils because they naturally adapt to high salinity, improving soil health over time without the high costs of physical or chemical interventions. By stabilizing soil structure, reducing erosion, and enhancing nutrient cycling, these cultivars sustainably improve both soil quality and agricultural productivity [15] [16].

Principal Component Analysis (PCA) is a widely used statistical technique in plant breeding that simplifies data analysis, identifies patterns, and guides decision-making [17]. Plant breeding to evaluate stress tolerance involves assessing numerous traits, including yield, quality, as well as morphological and physiological characteristics, which makes data management a complex task. PCA addresses this by reducing the dimensionality of datasets while retaining most of the variation, thereby simplifying interpretation. It groups correlated traits into principal components (PCs), helping breeders understand relationships among traits. For instance, traits like root length and yield might load heavily on the same component, indicating a strong association. This grouping also allows breeders to identify key traits contributing most to variation, enabling them to focus on the most impactful traits for selection [18].

PCA is essential for distinguishing between genotypes based on phenotypic or genotypic data, often through scatter plots that visually separate genotypes, aiding the identification of unique or promising candidates. It also plays a critical role in designing breeding programs by guiding the selection of parents with complementary traits, ensuring effective combinations [18][19]. Additionally, PCA is valuable in stress and adaptation studies, where it evaluates plant performance under different conditions by grouping traits or environments based on responses. This facilitates the selection of genotypes with broad adaptability or specific stress tolerance [20]. Its visualization capabilities in 2D or 3D plots enhance the intuitive understanding of complex datasets.

Principal Component Analysis is a powerful statistical tool used to study salt stress tolerance in plants by identifying and summarizing key variables that contribute to phenotypic or physiological variations. Under salt stress conditions, plants exhibit complex responses involving multiple traits, such as growth rate, photosynthetic efficiency and osmotic adjustments. PCA helps reduce the dimensionality of such datasets by grouping correlated traits into principal components (PCs), each representing a distinct aspect of variation. By identifying these associations, one can get insights into which traits contribute to enhanced salt tolerance, which could guide breeding programs focused on improving resilience in kodo millet.

2. MATERIAL AND METHODS

The experiment was executed in a polyhouse at the Research Farm of the Department of Genetics & Plant Breeding, College of Agriculture, RVSKVV, Gwalior, MP, India. The study material consisted of 42 kodo millet genotypes, with 40 sourced from ICRISAT, Hyderabad, and two varieties, JK 137 and JK 155, obtained from Jawaharlal Nehru Krishi Vishwavidalaya (JNKVV), Jabalpur, MP, India. A even growth medium of compost, vermiculite, and cocopeat in a 1:1:1 ratio was prepared to maintain consistency and avoid weed problem. The setup comprised nine plastic germination trays, each containing 112 cells with a height of 4 cm. Completely Randomized Design (CRD) was used with three replications across both control and stress conditions (NaCl concentrations of 50 mM, 100 mM, 250 mM, and 500 mM). Seeds were planted at a depth of 1 cm, with four seeds per cell. Salt stress tolerance was assessed by irrigating with specified NaCl concentrations from the first irrigation, while the control group received distilled water. Observations of morpho-physiochemical traits were logged 30 days post-germination. The ten morphological traits analyzed included days of emergence, final germination percentage, shoot length, root length, seedling length, shoot fresh weight, root fresh weight, total fresh weight, total dry weight, and vigor index. Physiochemical traits measured were proline content, chlorophyll a, chlorophyll b, total chlorophyll concentration, chlorophyll index, and salt tolerance index. Principal Component Analysis was performed using xlstat software.

Table 1: List of kodo millet genotypes used in the investigation

S.No.	Genotypes	S.No.	Genotypes	S.No.	Genotypes	S.No.	Genotypes
1.	IPs 4	**12.**	IPs 429	**23.**	IPs 706	**34.**	IPs 862
2.	IPs 5	**13.**	IPs 585	**24.**	IPs 730	**35.**	IPs 870
3.	IPs 91	**14.**	IPs 606	**25.**	IPs 741	**36.**	IPs 883
4.	IPs 105	**15.**	IPs 627	**26.**	IPs 764	**37.**	IPs 891
5.	IPs 176	**16.**	IPs 628	**27.**	IPs 777	**38.**	IPs 908
6.	IPs 181	**17.**	IPs 653	**28.**	IPs 782	**39.**	IPs 919
7.	IPs 208	**18.**	IPs 654	**29.**	IPs 785	**40.**	IPs 928
8.	IPs 287	**19.**	IPs 670	**30.**	IPs 793	**41.**	JK 137
9.	IPs 319	**20.**	IPs 694	**31.**	IPs 795	**42.**	JK 155
10.	IPs 358	**21.**	IPs 695	**32.**	IPs 814		
11.	IPs 388	**22.**	IPs 699	**33.**	IPs 828		

3. RESULTS AND DISCUSSION

At concentrations greater than 50 mM (namely 100 mM, 250 mM, and 500 mM), the impacts were intense, leading to total mortality and no germination. As a result, all observations were conducted for salt stress conditions solely at a concentration of 50 mM NaCl.

Out of the 12 principal components (PCs) analyzed, only two exhibited eigen values greater than 1, collectively accounting for 72.2% of the total variability among the studied traits. PC1 contributed the most to the variability, explaining 63.05% variation, followed by PC2, which accounted for 9.15% of the variation (Table 1). The eigen value of a component is the variance of the data explained by that component. A higher eigen value indicates greater importance of the component. If eigen value greater than 1 the component captures more variance than any single original variable and it contributes significantly in explaining the data variability. If eigen value is lesser than 1 the component captures less variance than a single variable, so it is usually discarded [21] [22]. Here scree plot is used as a visual tool to assess the importance of each principal component. It represents the proportion of the total variance explained by each principal component in descending order (Figure 1).

Table 2: Eigen values, variability percentage and cumulative variability percentage of 12 principal components

Particulars	**PC1**	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**	**PC11**	**PC12**
Eigen value	10.08	1.46	0.89	0.81	0.68	0.53	0.49	0.32	0.27	0.19	0.12	0.09
Variability (%)	63.05	9.15	5.61	5.10	4.26	3.35	3.10	2.01	1.71	1.18	0.75	0.60
Cumulative variability (%)	63.05	72.21	77.83	82.93	87.19	90.54	93.65	95.67	97.38	98.56	99.32	99.93

Table 2 reveals that each principal component is separately loaded with various morphological and physiochemical traits under investigation known as factor loadings. Value for factor loading revealed that the first principal component (PC1) which accounted for the highest variation was related with traits *viz.,* days of emergence, final germination percentage, shoot length, seedling length, shoot fresh weight, root fresh weight, total fresh weight, total dry weight and proline. The second principal component (PC2) was dominated by traits *viz.*, root length, vigor index, chlorophyll a and chlorophyll b. Factor loadings represent the relationship between the original variables and the principal components. These values indicate how strongly each variable contributes to or correlates with a specific principal component. Factor loadings range from -1 to 1, where values close to 1 or -1 signify a strong positive or negative correlation, respectively, and values near 0 indicate little to no relationship with the component. They are crucial for understanding the structure of the data because they describe the weight of each variable in forming a principal component [23] [24][25].

Table 3: Factor loadings for traits under study

Characters	PC1	PC2
Days of emergence (days)	**0.896**	0.185
Final germination percentage (%)	**0.795**	-0.206
Shoot length (cm)	**0.901**	-0.158
Root length (cm)	0.156	**0.529**
Seedling length (cm)	**0.918**	-0.161
Shoot fresh weight (g)	**0.839**	-0.266
Root fresh weight (g)	**0.742**	-0.046
Total fresh weight (g)	**0.873**	-0.219
Total dry weight (g)	**0.599**	-0.291
Vigor index	-0.421	**0.775**
Proline (µg g−1 fresh weight)	**0.667**	0.417
Chlorophyll a (mg g−1 tissue fresh weight)	0.509	**0.739**
Chlorophyll b (mg g−1 tissue fresh weight)	0.430	**0.566**
Total chlorophyll (mg g−1 tissue fresh weight)	0.785	0.548
Chlorophyll index	0.839	0.121
Salt tolerance index (%)	0.776	-0.063

The factor score of the each component (PC1 and PC2) had positive and negative values. Factor scores greater than 1 are highlighted in bold in Table 3. Factor scores represent the transformed values of the original data points in the reduced dimensional space defined by the principal components. In PC1, the positive scores ranged from 0.002 (Ips 670) to 3.795 (Ips 628), while negative value ranged from -7.521 (Ips 870) to -0.034 (Ips 764). In PC2, the positive value of the component arrayed from 0.061(Ips 795) to 2.894 (**Ips 5)** and negative value ranged from -2.687 (Ips 695) to -0.089 (Ips 764). Studying factor scores in principal component analysis is crucial as they represent individual genotypes in the reduced-dimensional space [18]. Derived from the linear combination of original variables weighted by component loadings, factor scores help reveal patterns, groupings, and outliers in the dataset, enabling a clearer understanding of the distribution and clustering of data based on the most significant variance dimensions [23] [26].

Genotypes which are common in PC1 and PC2 are Ips 5, Ips 208 and Ips 706 indicating that these genotypes contribute significantly to the variability captured by multiple principal components and thus for all the principal components under study. This suggests that these genotypes are influential and well-represented in the reduced-dimensional space, as they align strongly with the

dominant patterns or structures in the dataset [6] [27]. Such genotypes may exhibit attributes that are central to the underlying relationships among variables, reflecting a strong, consistent presence in the most important dimensions of variance. Identifying these genotypes can help highlight key patterns or representative observations in the data.

The biplot (Figure 2) typically displays the first two principal components (PC1 and PC2) as the axes, as these components capture the most of the variability. Each data point (observation) is plotted in the reduced-dimensional space using its factor scores. Variables in a biplot are represented as vectors (arrows) radiating from the origin [27] [28]. The direction of the arrows indicates how each variable contributes to the principal components, while the length reflects the strength of this influence. Longer arrows signify a stronger contribution of the variable to the principal components, highlighting its importance in explaining the variance within the data. The angle between arrows in a biplot indicates the correlations between variables. Small angles suggest a high positive correlation, perpendicular arrows imply no correlation, and opposite directions signify a negative correlation. Additionally, the projection of observations onto the arrows represents how strongly an observation is associated with a particular variable, providing insight into the relationships between observations and variables [29] [30] [31].

4. CONCLUSION

Kodo millet, characterized by small floret size and a high degree of self-pollination, poses significant challenges for hybridization. Consequently, pure line and mass selection from high-yielding germplasm accessions are often preferred methods for varietal development. In India, most small millet varieties have been developed from existing germplasm. Evaluating preserved germplasm for salt stress tolerance offers significant potential to enhance Kodo millet yields [32]. Principal Component Analysis (PCA) is a valuable tool in both pure line and mass selection, as it simplifies the assessment of genetic variability and trait significance. In pure line selection, PCA identifies traits contributing to variability, clusters similar lines, and highlights those performing well under specific conditions, such as stress. In mass selection, PCA evaluates population diversity, identifies critical traits, tracks selection progress, and examines environmental interactions. By reducing data complexity and focusing on traits driving variability, PCA improves the efficiency and accuracy of breeding strategies in both approaches.

.

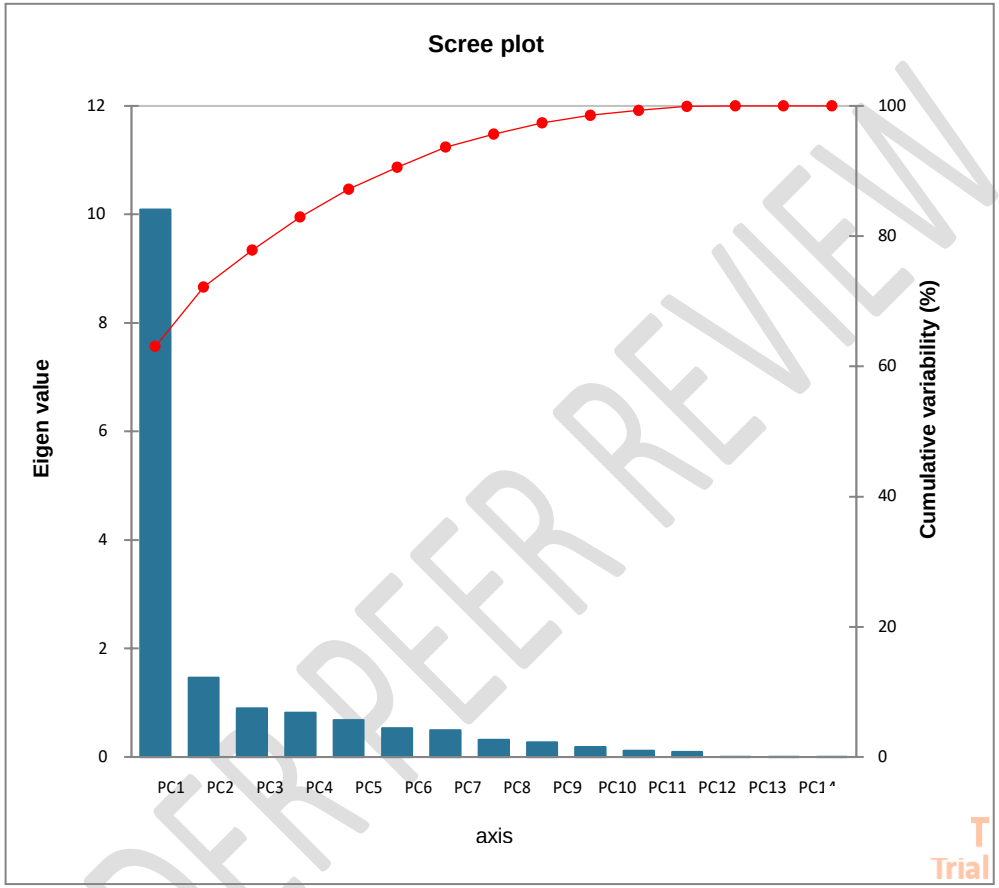

FIG.1. Scree plot between Eigen value and Principal components

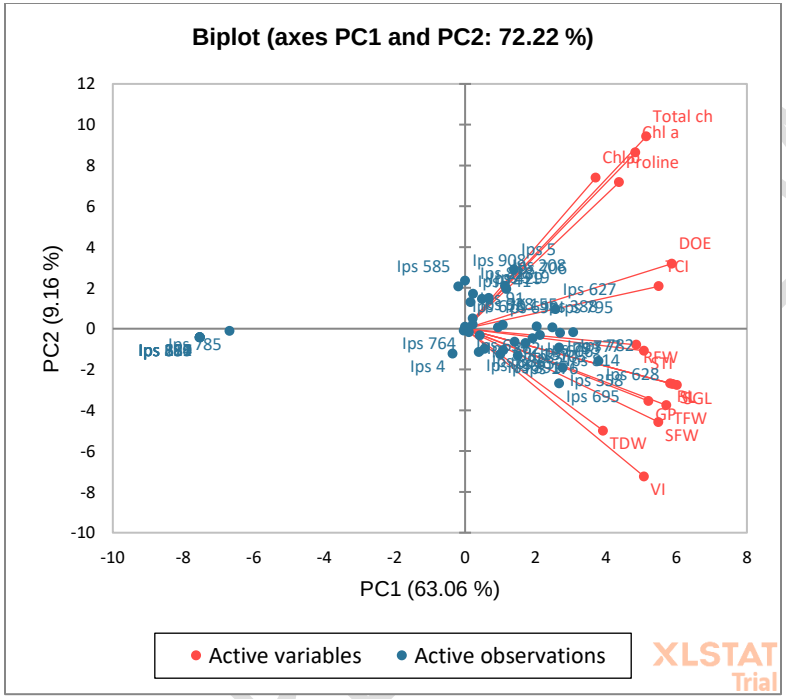

FIG.2. Biplot against PC1 and PC2 for studied characters of 42 genotypes

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

REFERENCES

1.Gyawali, P. (2021). Production trend, constraints, and strategies for millet cultivation in Nepal: A study from review perspective. *International Journal of agricultural and applied sciences*, *2*(1), 30-40.
DOI: https://doi.org/10.52804/ijaas2021.213
2.Swapnil, Parveen, R., Singh, D., Imam, Z., & Singh, M. K. (2024). Breeding Kodo Millet for Biotic and Abiotic Stress Tolerance. In *Genetic improvement of Small Millets* (pp. 613-635). Singapore: Springer Nature Singapore.
DOI: https://doi.org/10.1007/978-981-99-7232-6_30
3.Ravikesavan, R., Jeeva, G., Jency, J. P., Muthamilarasan, M., & Francis, N. (2023). Kodo millet (Paspalum scorbiculatum L.). In *Neglected and Underutilized Crops* (pp. 279-304). academic press.
DOI: https://doi.org/10.1016/B978-0-323-90537-4.00019-3
4.Sanodiya, L. K., Sanodiya, P. K., Prajapati, A., Kumawat, P., & Singh, A. (2023). The sustainability of millets and the advantages they bring for human health. *Research Gate*.
DOI: https://doi.org/10.5281/zenodo.7947530
5.Deshpande, S. S., Mohapatra, D., Tripathi, M. K., & Sadvatha, R. H. (2015). Kodo millet-nutritional value and utilization in Indian foods. *Journal of grain processing and storage*, *2*(2), 16-23.
6. Vetriventhan, M., & Upadhyaya, H. D. (2019). Variability for productivity and nutritional traits in germplasm of kodo millet, an underutilized nutrient-rich climate smart crop. *Crop Science*, *59*(3), 1095-1106.
DOI: https://doi.org/10.2135/cropsci2018.07.0450
7.Arya, S. S., & Shakya, N. K. (2021). High fiber, low glycaemic index (GI) prebiotic multigrain functional beverage from barnyard, foxtail and kodo millet. *Lwt*, *135*, 109991.
DOI: https://doi.org/10.1016/j.lwt.2020.109991
8. Bunkar, D. S., Goyal, S. K., Meena, K. K., & Kamalvanshi, V. (2021). Nutritional, functional role of kodo millet and its processing: a review. *International Journal of Current Microbiology and Applied Sciences*, *10*(01), 1972-1985.
DOI: https://doi.org/10.20546/ijcmas.2021.1001.229
9. Priya, K., & Gakkhar, A. Nutritional Importance of Miillet in India.
10.Yadav, M., Bansal, C., Shukla, U., & Zainab, S. (2024). A critical review on the efficacy of Kodo, Sorghum, and foxtail millets in the management of type 2 diabetes mellitus. *Indian Journal of Health Sciences and Biomedical Research kleu*, *17*(3), 200-208.
DOI: 10.4103/kleuhsj.kleuhsj_243_24
11. Flowers TJ, Colmer TD. Salinity tolerance in halophytes. New phytologist. 2008 Sep 1:945-63.
DOI: 10.1111/j.1469-8137.2008.02531.x
12. Shabala S. Learning from halophytes: physiological basis and strategies to improve abiotic stress tolerance in crops. Annals of botany. 2013 Nov 1;112(7):1209-21.

DOI: https://doi.org/10.1093/aob/mct205
13.Kumar, R., Singh, A., Bhardwaj, A. K., Kumar, A., Yadav, R. K., & Sharma, P. C. (2022). Reclamation of salt-affected soils in India: Progress, emerging challenges, and future strategies. *Land Degradation & Development*, *33*(13), 2169-2180.
DOI: https://doi.org/10.1002/ldr.4320
14.Ashraf, M., & Munns, R. (2022). Evolution of approaches to increase the salt tolerance of crops. *Critical Reviews in Plant Sciences*, *41*(2), 128-160.
DOI: https://doi.org/10.1080/07352689.2022.2065136
15.Ashraf, M., Athar, H. R., Harris, P. J. C., & Kwon, T. R. (2008). Some prospective strategies for improving crop salt tolerance. *Advances in agronomy*, *97*, 45-110.
DOI: https://doi.org/10.1016/S0065-2113(07)00002-8
16.Munns, R., & Gilliham, M. (2015). Salinity tolerance of crops–what is the cost?. *New phytologist*, *208*(3), 668-673.
17. Khatun, M., Islam, M. R., Khan, M. A. R., Hossain, M. K., & Islam, A. K. M. A. (2023). Principal component analysis and estimated breeding values for selecting suitable parental genotypes in rice (Oryza sativa L.). *Sabrao Journal of Breeding & Genetics*, *55*(6).
DOI: http://doi.org/10.54910/sabrao2023.55.6.3
18. Roka, P., Shrestha, S., Adhikari, S. P., Neupane, A., Shreepaili, B., & Bista, M. K. (2024). A review on genetic parameters estimation, trait association, and multivariate analysis for crop improvement. *Archives of Agriculture and Environmental Science*, *9*(3), 618-625.
DOI: https://doi.org/10.26832/24566632.2024.0903029
19. Mehta, A., Sengupta, P., Garg, D., Singh, H., & Diamand, Y. S. (2023). Benchmarking the Effectiveness of Classification Algorithms and SVM Kernels for Dry Beans. *arXiv preprint arXiv:2307.07863*.
DOI: https://doi.org/10.48550/arXiv.2307.07863
20. Füzy, A., Kovács, R., Cseresnyés, I., Parádi, I., Szili-Kovács, T., Kelemen, B., ... & Takács, T. (2019). Selection of plant physiological parameters to detect stress effects in pot experiments using principal component analysis. *Acta Physiologiae Plantarum*, *41*, 1-10.
DOI: https://doi.org/10.1007/s11738-019-2842-9
21. Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, *29*(2), 295-327.
DOI: 10.1214/aos/1009210544
22. Danisman, Y., Yilmaz, M. F., Ozkaya, A., & Comlekciler, I. (2014). A comparison of eigenvalue methods for principal component analysis. *Appl. and Comput. Math*, *13*, 316-331.
23. Jolliffe, I. T. (2002). Principal component analysis and factor analysis. *Principal component analysis*, 150-166.
DOI: https://doi.org/10.1007/0-387-22440-8_7
24. De Winter, J. C., & Dodou, D. (2016). Common factor analysis versus principal component analysis: a comparison of loadings by means of simulations. *Communications in Statistics-Simulation and Computation*, *45*(1), 299-321.
25. Yamamoto, H., Fujimori, T., Sato, H., Ishikawa, G., Kami, K., & Ohashi, Y. (2014). Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis. *BMC bioinformatics*, *15*, 1-9.
DOI:10.1186/1471-2105-15-51
26. Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433-459.
DOI: https://doi.org/10.1002/wics.101
27. Shashibhushan, D., Kumar, C. S., & Kondi, R. K. R. (2022). Principal Component Analysis for Yield and Yield Related Traits in Pearl Millet Cultivars. *Ecol. Environ. Conserv*, *28*, S104-S107.

DOI: http://doi.org/10.53550/EEC.2022.v28i06s.017
28. Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*(3), 453-467
DOI: https://doi.org/10.1093/biomet/58.3.453
29. Wedlake, R. S. (2008). *Robust principal component analysis biplots* (Doctoral dissertation, Stellenbosch: University of Stellenbosch.
30. Gower, J. C., & Hand, D. J. (1995). *Biplots* (Vol. 54). CRC Press.
31. Holland, S. M. (2008). Principal components analysis (PCA). *Department of Geology, University of Georgia, Athens, GA*, *30602*, 2501.
32. SAO, A., PALI, V., & PATIL, H. E. (2024). Genetic improvement for yield, quality, biotic, and abiotic stresses in little millet (panicum sumatrense roth. ex roem. and schult.). in *genetic improvement of small millets* (pp. 571-599). singapore: springer nature singapore.
https://DOI.ORG/10.1007/978-981-99-7232-6_28