Electronic Journal of Plant Breeding

Research Article

Linear discriminant analysis to differentiate sorghum germplasm: A crucial tool for breeding programmes

Neladri Sekhar Sarkar[1], T. Kalaimagal[1]*, D. Kavithamani[1], R. Chandirakala[1], S. Manonmani[1] M. Raveendran[2] and A. Senthil[3]

[1]Centre for Plant Breeding and Genetics, TNAU, Coimbatore, India
[2]Directorate of Research, Dept. of Plant Biotechnology, CPMB&B, TNAU, Coimbatore
[3]Department of Crop Physiology, TNAU, Coimbatore-641003
*E-Mail: kalaimagal.t@tnau.ac.in

Abstract
A total of 86 Sorghum genotypes along with three checks (CO 30, CO 32, and K 12) were evaluated during *Rabi* 2021 season to identify variations and character associations among grain yield and yield component traits. The phenotypic data collected were used to create a statistical database and were analyzed using linear discriminant analysis (LDA) to identify and discriminate landraces for utilization in sorghum breeding. The LDA successfully differentiated the genotypes into three groups with an accuracy of 73.52%. The study revealed a significant level of variation among the genotypes, based on observations for nine quantitative traits. Further analysis using the LDA biplot showed that the genotypes within clusters 1 and 4 hold potential for future breeding programmes. Therefore, the observed phenotypic data can be useful for identifying and selecting appropriate accessions for sorghum improvement.

Keywords : Sorghum, Landraces, Clusters, Linear discriminate analysis(LDA)

INTRODUCTION

Sorghum (*Sorghum bicolor*) is a cereal crop that belongs to *Poaceae* family, a diverse group comprising nearly 12,000 species, including several significant subfamilies. Within this family, the *Sorghum* genus encompasses 31 cultivated species and 17 related wild species (Snowden, 1936). It is difficult to track sorghum domestication due to lack of archaeological shreds of evidence (de Wet & Huckabay, 1967). However, a few reports suggested that at the Egyptian Sudanese border the domestication might have happened around 5000–8000 years ago (Mann *et al*., 1983). It is one of the most important staple food for Africa and Asian countries. Globally, it is the fifth most important cereal crop after wheat, rice, maize, and barley (Cuevas *et al*., 2014).

Sorghum is a versatile crop grown under adverse climatic conditions ranging from arid and semi-arid regions to tropical and temperate areas, and used for animal feed, biofuel, and various industrial products besides human consumption. One of its notable characteristics is the deep root system that allows efficient absorption of moisture and nutrients from the soil. Sorghum's resilience is evident in its ability to withstand drought, high temperatures, and pests. Sorghum exhibits successful growth in a wide range of soil types and is adaptable to diverse temperature conditions. (Nguyen *et al*., 2013).

Sorghum plays an important role as nutritional security crop for its nutritional quality and ability to grow under adverse environmental conditions (Hariprasanna and Patil, 2015). India contributes approximately 16% of global sorghum production (Rao and Parwez, 2003), and it contains significant amount of carbohydrate, protein, fibre, iron, calcium, and phosphorus. In addition,

sorghum's gluten-free nature has made it an attractive substitute for wheat with several health benefits besides, reducing cholesterol levels, inflammation etc.

The statistical approach, Linear Discriminant Analysis (LDA) in sorghum breeding is used to differentiate and classify diverse genotypes based on various morphological, physiological, and molecular characteristics as observed by Boedeker and Kearns (2019). By analyzing these characteristics, LDA provides valuable insights into the genetic factors underlying observed variations. This information is crucial for breeders and geneticists for identification and selection of promising genotypes for further improvement. A notable advantage of LDA is its ability to handle high-dimensional datasets by reducing dimensionality while maximizing the differentiation between groups.

MATERIALS AND METHODS
The experimental materials comprised of 86 sorghum germplasm obtained from Dr. Ramaiah gene bank of the Department of Plant Genetic Resources and the Department of Millets, TNAU, Coimbatore . These were evaluated during *Rabi* 2021-22 at Millet Breeding Station farm, TNAU. The experiment was raised in augmented block design I (ABD I) and each genotype was sown in

two rows with a row length of 4 meters at a spacing of 45cm x 15cm. The data on nine quantitative traits viz., plant height (PH), number of leaves (NOL), stem diameter (SD), panicle length (PL), panicle width (PW), panicle weight (PWt.), hundred seed weight (HSW), dry fodder yield per plant (DFY) and grain yield per plant (GYPP) were recorded for the genotypes including the accessions and the checks based on descriptor guidelines published by IBPGR/ICRISAT,1993. Observations were recorded on five randomly selected plants from each genotype and was discriminated by the Linear Discriminant Analysis (LDA) procedure proposed by Fisher, R. A. (1936) to identify and distinguish sorghum germplasm accessions suitable for sorghum breeding programmes. The statistical analysis was done using "psych", "devtools", "MASS", "ggord", "Klar", and "caret" packages in RStudio v 4.1.1.

RESULTS AND DISCUSSION
The use of Linear Discriminant Analysis showcased a remarkable level of discrimination (Sau *et al.,* 2018). The scatter plot was generated using various combinations of independent variables, revealing a greater degree of overlap among the groups (**Fig. 1**). The histogram predicted the frequency distribution of each variable. The correlation coefficient indicated that, most of the traits showed a significantly positive relationship with grain yield
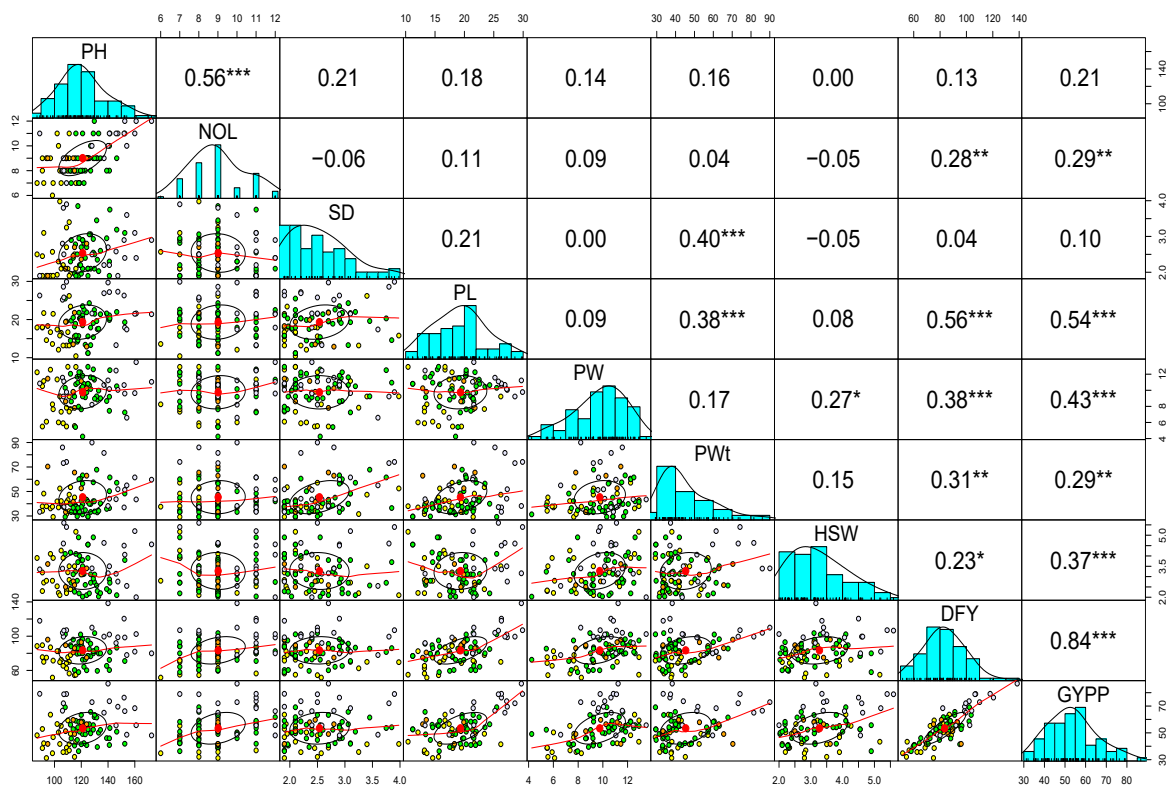

**Fig. 1. Scatter plot, Frequency distribution and Correlation coefficient of the variables**

Table 1. Germplasm accessions grouped into four clusters with hierarchical cluster analysis (HCA)

Clusters	Number of accessions	Accessions
Cluster I	63	IS 10370, SPV 615, YT 82164, IS 4646, MR 32B, AS 3890, B 35, CSV 4, RS 1804, A 6460, SB 79B, ISVAT 1016, AS 3863, IL 231, MR 8, A 3895, ICSV 166, SB 103, MR 119, A 404, IL 497, ICSB 29, MS 8901, IS 3962, IS 20575, IL 1136, AS 209, CSV 20, AS 3749, ICSV 137, AS 3883, E 16, AS 2157, E 3, SPV 386, A 3822, PM 7439, DM 153, DKV 3, ICSB 23, AS 2578, SAR 17.00, A 524, AS 2487, IL 594, MR 115, IL 528, IL 406, MR 76, ICSV 241, SAR 34.00, IL 98, AS 687, MR 47 , IS 18323, PC 53, IS 19539, IS 9283, IS 9652, COS 28, TNS 599, IL 4664 and AS 219
Cluster II	6	ICSPIR MER, ICSB 42, ICSV 61, MR 87, ICSPIR 132 and ICSV 202
Cluster III	8	IL 268, M106, M 35544, MR 119C, ICSB 31, AS 2586, MR 852 and AS 3816
Cluster IV	9	ICSP 28 MFR, MR 22/1, M 26405, MR 99, MR 77, AS 512, A 6072, ICSV 209 and E 19

Table 2. Coefficients of linear discriminants function of the nine traits

Variables	LD1	LD2	LD3
PH	-0.017648	-0.0690452	-0.0084624
NOL	0.284862	-0.1982506	0.22353332
SD	0.474355	0.32892834	-0.8223586
PL	0.15775686	0.0088442	-0.0417632
PW	-0.0807429	0.03223653	-0.0032941
PWt.	0.01591442	-0.0229767	0.03991588
HSW	0.76202841	-0.1847319	0.16811509
DFY	0.03468704	0.00574055	0.09173588
GYPP	0.01293346	0.01592985	-0.1207104

PH-Plant height, NOL-Number of leaves, SD-Stem diameter, PL-Panicle length , PW-Panicle width, PWt.-Panicle weight, HSW-Hundred seed weight, DFY-Dry fodder yield per plant, GYPP-grain yield per plant.
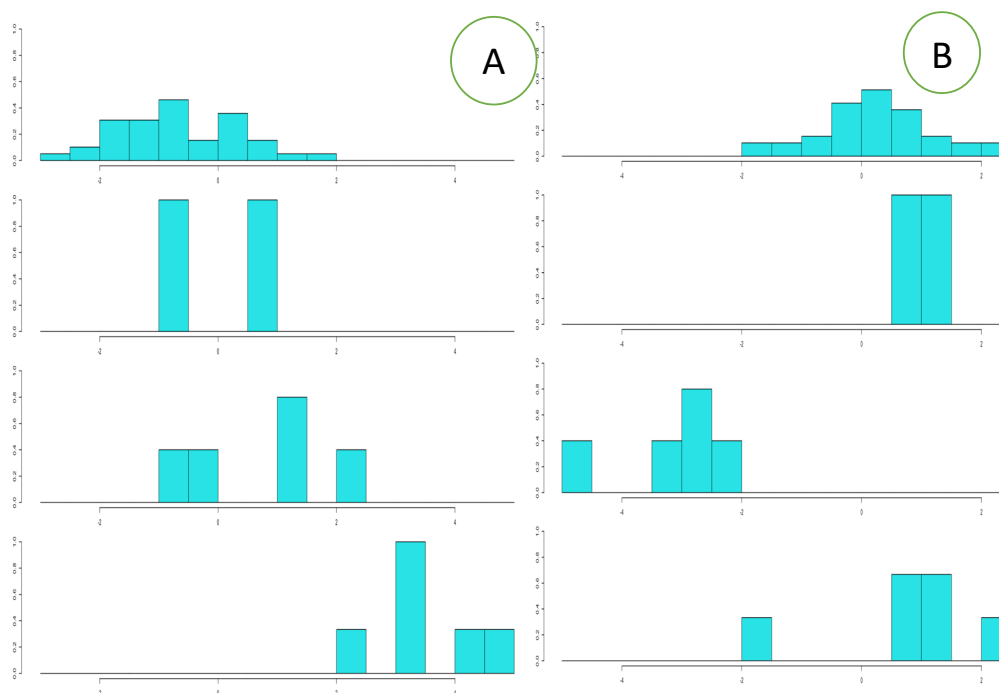
Fig. 2. Histogram for percentage of separation, (A) based on LD1, (B) based on LD2. The y-axis represents the frequency (or count) and the groups are taken along X-axis

per plant. Similar findings were reported by Sahib and Reddy (1990), Seetharam and Ganesamurthy (2013), Patil *et al.* (2014), Eudalamaw *et al.* (2017), Deshmukh *et al.* (2018) and Swamy *et al.* (2018), Jankovic *et al.* (2012), Jain and Patel (2014) and Shivaprasad *et al.* (2019). The proportion of samples in each class was defined as prior probability and calculated for each group (**Table 1**). It indicated that the accessions belonged to one of the clusters in between cluster 1 to 4 respectively. In the present study, the sorghum germplasm accessions were

categorized into four distinct groups based on clustering, with prior probabilities indicating that 75%, 3.8%, 9.6% and 11% of the genotypes belonged to clusters 1, 2, 3, and 4, respectively. The relative importance of predictors was calculated for each dependent variable as coefficients of linear discriminants (**Table 2**). The percentage of separation achieved by LD1 was 57.11%, by LD2 was 34.39%. Similar result were reported by Chinnasamy *et al.,* 2021 and very less amount of separation achieved by LD3 was about 8.49%.
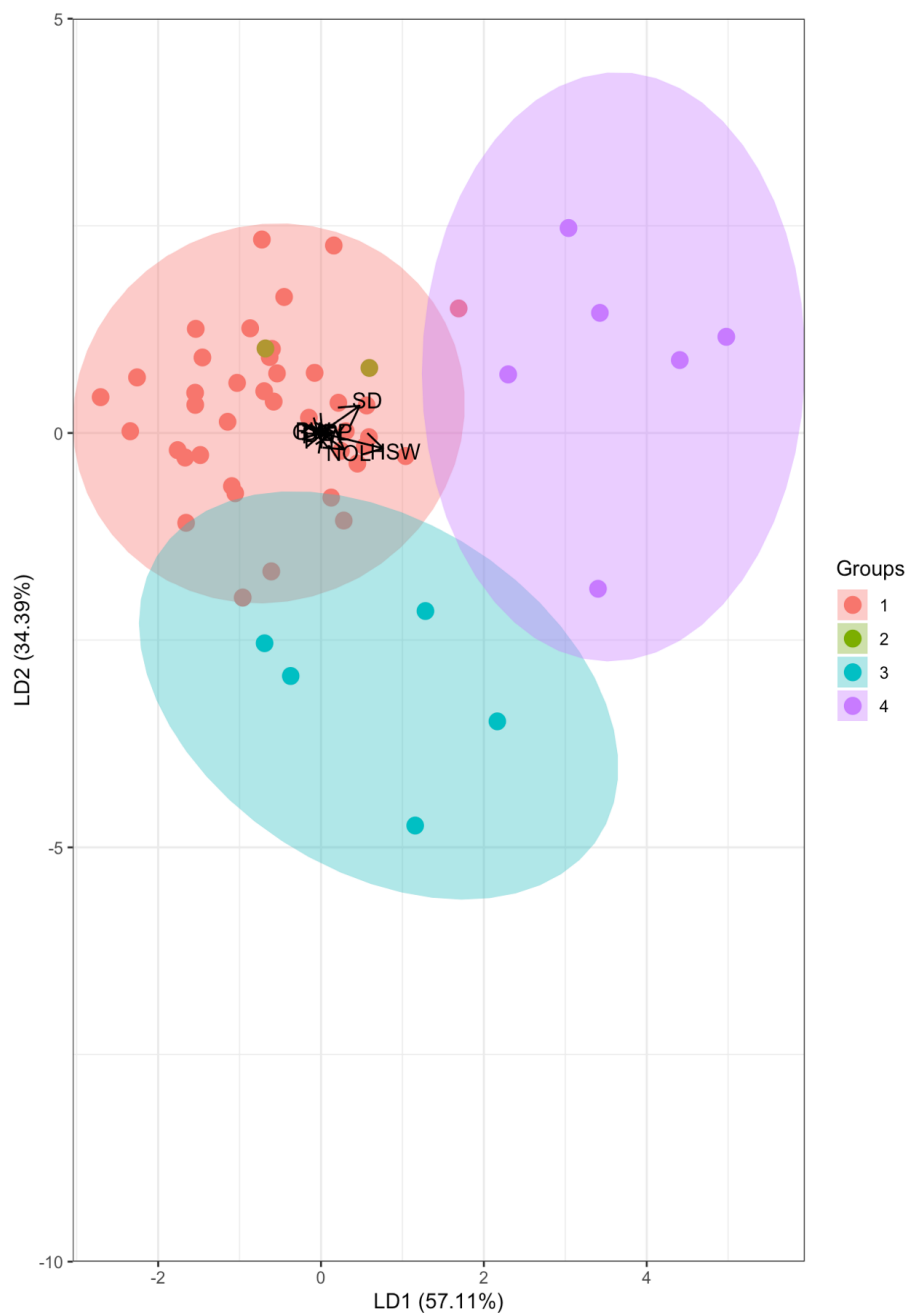

Fig. 3. Biplot generated between LD1 and LD2 of the clusters

The examination of the LD1 histogram demonstrated a clear distinction between clusters 1 and 4, with minimal instances of overlap. The X axis in the **fig. 2** showed the range of the clusters. The overlapping of groups was primarily due to duplication of genotypes from different groups. However, there was some noticeable overlap between clusters 1, 2, 3, and 4 with a separation of 57%, as depicted in **Fig. 2**. In contrast, the LD2 histogram showed a higher degree of overlap, with only 34% separation, which is significantly lower than the separation achieved by LD1.

The biplot generated in **Fig. 3** utilized LD1 and LD2 as the X and Y axes, respectively. This biplot effectively illustrated the separation of different groups, as indicated by the high separation percentages achieved by LD1 (57.11%) and LD2 (34.39%). It was seen that there was some overlap between all the clusters . In the biplot, cluster 1 ranged from -10 to +2, cluster 2 ranged from -1 to +1, cluster 3 ranged from -2 to +3, and cluster 4 ranged from +1 to +6. Clusters 2 and clusters 3 were situated in the middle. Notably, the genotypes in cluster 1 and cluster 4 exhibited greater diversity, as they were positioned further apart in the biplot.

The partition plot in **Fig. 4** was generated assuming a shared covariance matrix for all classes in the LDA. By examining the partition plot, it was observed that the classification of each observation in the training dataset using the Linear Discriminant Model, considering every combination of two variables. In the partition plot, it was evident that the genotypes belonging to cluster 4 are notably distinct and well-separated from the other clusters across most variable combinations. However, clusters 1, 2, and 3 displayed some degree of overlap, suggesting a relationship between these clusters. The study demonstrated that the method used to discriminate between genotypes was valid and achieved a high level of accuracy, (Alajas *et al.,* 2021). Quadratic Discriminant Analysis (QDA) partition plot is not possible here because majority of clusters were having very less number of genotypes. So, The **Table 1** represents the discriminated genotypes, suggesting that genotypes from clusters 1 and 4 hold potential for utilization in future breeding programmes. So, the cluster analysis revealed the diverse genotypes that could be used in hybridization programmes for exploiting the maximum heterotic potential (Subramanian *et al*., 2019). The accuracy percentage, which is calculated by dividing the predicted
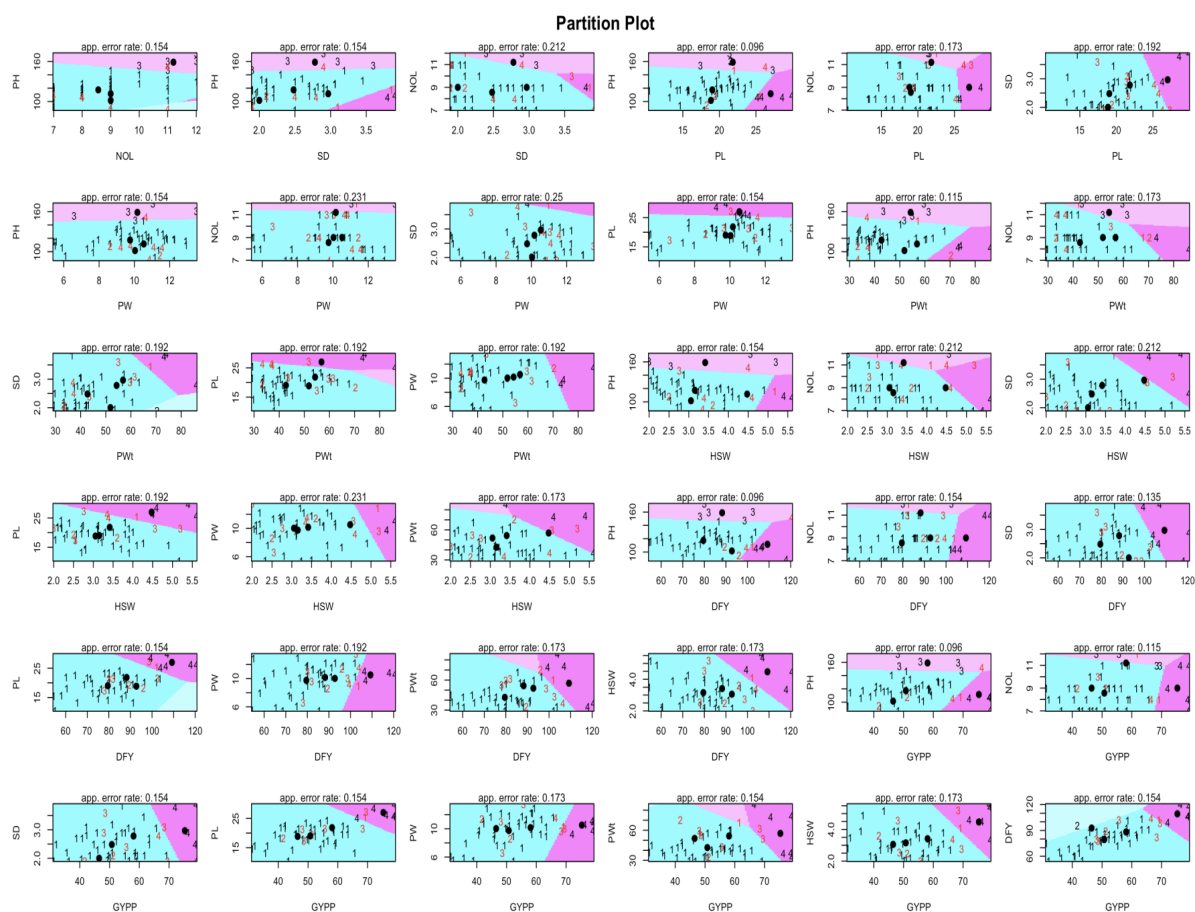

Fig. 4. Partition plot based on linear discriminate analysis

value by the actual value, is used to evaluate the accuracy of the discrimination method. The accuracy percentage was 73.52% (Huang *et al.,* 2019). The accuracy level achieved in the study was moderately good, which can be attributed to the minimal difference between the predicted values and the actual values. Further, the division of the linear discriminant function into four levels might have contributed to the high accuracy percentage observed.

Thus the analysis conducted on 86 Sorghum germplasm using LDA revealed a significant level of discrimination based on clusters. Clusters that exhibited complete separation from one another can be considered highly diverse and can be utilized in breeding programmes aimed at developing high-yielding and resilient grain varieties.

REFERENCES

Alajas, O.J., Concepcion, R., Dadios, E., Sybingco, E., Mendigoria, C.H. and Aquino, H. 2021. Prediction of Grape Leaf Black Rot Damaged Surface Percentage Using Hybrid Linear Discriminant Analysis and Decision Tree. International Conference on Intelligent Technologies (CONIT), pp. 1-6. IEEE. [Cross Ref]

Boedeker, P. and Kearns, N.T. 2019. Linear discriminant analysis for prediction of group membership: A User Friendly Primer. Linear discriminant analysis for prediction of group membership: A User Friendly Primer. *Advances in Methods and Practices in Psychological Science*, **2(3)**:250–263. [Cross Ref]

Chinnasamy, K., Arumugam, Y., Jegadeesan, R. and Chockalingam, V. 2021. Linear discriminant analysis in red sorghum using artificial intelligence. *The Nucleus*, **64**:103-113. [Cross Ref]

Cuevas, H.E., Prom, L.K., Erpelding, J.E. and Brotons, V. 2014. Assessments of genetic diversity and anthracnose disease response among Zimbabwe sorghum germplasm. *Plant Breeding,* **133(2)**:234-242. [Cross Ref]

de Wet, J.M.J. and Huckabay, J.P. 1967. The origin of Sorghum bicolor. II. Distribution and domestication. *Evolution*, **21:**787-802. [Cross Ref]

Deshmukh, S.B., Bagade, A.B. and Choudhari, A.K. 2018. Correlation and path analysis studies among *rabi* sorghum (*Sorghum bicolor* (L.) Moench) mutants. *International Journal of Current Microbiology and Applied Sciences,* **6**: 1014-1020.

Eudalamaw, C., Adgna, A. and Mohammed, H. 2017. Correlation and path coefficient analysis of agronomic and quality traits in a bioenergy crop, sweet sorghum [*sorghum biocoior*(L.) Moench]. *African Journal of Biotechnology,* **16**(47): 2189-2200. [Cross Ref]

Fisher, R.A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* **7**(2):179–188. [Cross Ref]

Hariprasanna, K. and Patil, J.V. 2015. Sorghum: origin, classification, biology and improvement. *Sorghum Molecular Breeding*, **1**:3–20. [Cross Ref]

Huang, L., Wu, Z., Huang, W., Ma, H. and Zhao, J. 2019. Identification of fusarium head blight in winter wheat ears based on Fisher's linear discriminant analysis and a support vector machine. *Applied Sciences*, **9**(18):3894. [Cross Ref]

Jankovic, S., Rakic, S., Ikanovic, J., Kuzevski, J., Zivanovic, L. and Lakic, Z. 2012. Correlation coefficients of morphological - productive traits of species of Sorghum genus. *Biotechnology in Animal Husbandry,* **28**(3): 585-593. [Cross Ref]

Jain, S.K. and Patel, P.R. 2014. Characters association and path analysis in sorghum [*sorghum bicolor*(L.) Moench] f$_1$s and their parents, *Annals of Plant and Soil Research,* **16**(2): 107-110.

Mann, J.A., Kimber, C.T. and Miller, F.R. 1983. The origin and early cultivation of sorghums in Africa. Bulletin (Texas agricultural experiment station), 1454. College Station: Texas Agricultural Experiment Station.

Nguyen, C.T., Singh, V., van Oosterom, E.J., Chapman, S.C., Jordan, D.R. and Hammer, G.L. 2013. Genetic variability in high temperature effects on seed-set in sorghum. *Functional Plant Biology,* **40(5)**: 439-448. [Cross Ref]

Patil, C.N., Rathod, A.H., Vaghela, P.O., Yadav, S.R., Patade, S.S. and Shinde, A.S. 2014. Study of correlation and path analysis in dual purpose sorghum [*Sorghum bicolor*(L.) Moench]. *International Journal of Agricultural Sciences,* **10**(2): 608-611.

Rao, B.D. and Parwez, S. Md. 2003. Analysis of growth in sorghum production in India. *Asian Econ. Rev,* **45**(3): 503-515.

Sau, S., Ucchesu, M., Dondini, L., De Franceschi, P., D'hallewin, G. and Bacchetta, G. 2018. Seed morphometry is suitable for apple-germplasm diversity-analyses. *Computers and Electronics in Agriculture,* **151**:118–25. [Cross Ref]

Snowden, J.D. 1936. Cultivated races of sorghum. London: *Allard and Sons.* P. 274.

Seetharam, K. and Ganesamurthy, K. 2013. Characterization of sorghum genotypes for yield and other agronomic traits through genetic variability and diversity analysis. *Electronic Journal of Plant Breeding,* **4**(1)**:** 1073-1079.

Swamy, N., Biradar, B.D., Sajjanar, G.M., Ashwathama, V.H., Sajjan, A.S. and Biradar, A.P. 2018. Genetic variability and correlation studies for productivity traits in *Rabi* sorghum [*Sorghum bicolor*(L.) Moench]. *Journal of Pharmacognosy and Phytochemistry*, **7**(6): 1785-1788.

Shivaprasad, T., Girish, G., Badigannavar, A., Muniswamy, S., Yogesh, L.N. and Ganapathi, T.R. 2019. Genetic variability, correlation and path coefficient studies in sorghum [Sorghum bicolor (L.) Moench] mutants. *Electronic Journal of Plant Breeding*, **10** (4):1383-1389. [Cross Ref]

Subramanian, A., Raj, R.N. and Elangovan, M. 2019. Genetic variability and multivariate analysis in sorghum (Sorghum bicolour) under sodic soil conditions. *Electronic Journal of Plant Breeding*, **10** (4):1405-1414. [Cross Ref]

Sahib, K.H. and Reddy, B. 1990. Character Association And Path Analysis In *Rabi* Sorghum [Sorghum Bicolor (L.) Moench]. *Madras Agricultural Journal*, **77**(Sep-Dec):1. [Cross Ref]