PRINCIPAL COMPONENT ANALYSIS AND GROUPING OF BROWN MIDRIB SORGHUM (SORGHUM BICOLOR L. MOENCH) ADVANCED BREEDING LINES FOR GENETIC DIVERSITY

MANISH V BORATKAR*, HASANALI NADAF, ANIL REDDY AND PRABHAKAR BABU GUNTURU

Foragen Seeds Private Limited, Hyderabad-500034 (Telangana), India *(e-mail: mb.cotbreeder@gmail.com)

(Received: 03 June 2025; Accepted: 28 June 2025)

SUMMARY

Principal component and hierarchical cluster analyses were carried out with six quantitative characters of hundred and ninety four brown midrib sorghum advanced breeding lines. Principal factor analysis identified three principal components which explained about 62.5 % variability. PC 1 explained maximum variability of 24.7% of total variation for morphological characters and PC 2 loaded with 20.2% of total phenotypic variability and PC 3 had contributed 17.5 % of the total variation. PCA revealed that Days to 50% flowering, Plant height (cm), Internode length (cm) and Green fodder yield (t/ha) depending upon their loading variables on a common principal axis. Brown midrib Sorghum advanced breeding lines *viz.*, FBMRS-24, FBMRS-31, FBMRS-42, FBMRS-43, FBMRS-87, FBMRS-88 and FBMRS-167 were identified as superior lines based on principal component analysis. Hierarchical cluster analysis emphasized on 194 BMR sorghum advanced lines and grouped into twenty clusters. Cluster number 1 had only two lines and both were to be superior for highest green fodder yield and highest brix percentage. Both PCA and clustering registered same level of variability between breeding lines. Hence these superior brown midrib advanced breeding lines may further be utilized in breeding programmes for developing BMR sorghum varieties with high yield and for developing high yielding BMR hybrids.

Key words: Brown midrib, sorghum, green fodder yield and brix percentage

Sorghum [Sorghum bicolour (L.) Moench] is in the fifth position among cereal crop globally after wheat (Triticum aestivum), rice (Oryza sativa), maize (Zea mays) and barley (Hordeum vulgare). In India sorghum occupied 39.8 million hectare area, 48.8 lakh tons of production and 1225 kg per ha productivity (DA&FW, Third Advance Estimate 2024-25). The largest diversity of the crop germplasm provides greater opportunities for improvement regarding its environmental adaptability and acquiring better agronomic traits from the crop species. Identifying and selecting the best varieties meeting specific local food and industrial requirements from this great biodiversity is of high importance for the food security assurance of any given country. Having a good knowledge of the genetic diversity of a crop often enables the plant geneticist to select the desirable family for the breeding programme and gene introgression from distantly related germplasm.

Principal component analysis (PCA) is one of the best dimension reduction technique for reducing

the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. Finding such new variables, the principal components, reduces to solving an eigenvalue/eigenvector problem, and the new variables are defined by the dataset at hand, not a priori, hence making PCA an adaptive data analysis technique (Jolliffe and Cadima, 2016). The main objective of this research was to determine the range of variation among brown midrib advanced sorghum lines in general and to classify them into clusters based on their similarity features regarding the quantitative characters under study and also to generate data on their performance for plant breeders for further evaluation of the crop in particular.

MATERIALS AND METHODS

A total of one hundred and ninety four brown midrib advanced sorghum lines (F_7 generation) were planted for characterization and seed multiplication

(Table 1). Crop was raised during rabi 2024 at Foragen seeds R&D farm, Hyderabad. The soil for the experiment was sandy loam to loam with uniform topography and free from waterlogged conditions. Each genotype was grown in 1 row of 2 meter with a spacing of 60 cm between rows and 15 cm between plants. Optimum plant population was maintained. Standard agronomic package and practices were followed to raise a healthy crop. The analysis is based on the determination of the eight quantitative traits viz., days to 50% flowering, plant height (cm), Number of productive tillers, Internode length, refractometer brix percentage and green fodder yield (t/ha) at maturity. Analysis of variance using descriptive statistics such as mean, standard deviation and coefficient of variation for all the six traits were calculated. Clustering of genotypes into similar groups was performed. To identify the patterns of morphological variation, (PCA) was conducted. The observations recorded on six traits were statistically analysed by using GRAPES package for analysis.

RESULTS AND DISCUSSION

The First degree Statistical measures *viz.*, maximum, minimum, sum, mean, Standard Deviation (SD) and Coefficient of Variation (CV) for the measured traits are shown in Table 1. The largest variation was observed for number of productive tillers with CV of 0.39% followed by green fodder yield 0.1%%, refractometer brix recorded the variation of 0.20%, internode length recorded 0.12% and plant height 0.08% variation among different morphological traits studied. The minimum level of variation was observed by the trait and days to 50% flowering 0.04%.

The genotype FBMRS-14 has taken the minimum of 74 days for days to 50% flowering. Whereas, the longest duration of 88 days had taken by FBMRS-167. The overall mean for days to 50% flowering in brown midrib advanced sorghum lines was 77 days. The mean value for plant height was 178.03 cm with the minimum and maximum value of 148 cm and 225 cm were recorded by the genotypes FBMRS-162 and FBMRS-150 respectively. For greater coefficient of variation (0.39%) was observed for number of productive tillers i. e., eleven lines showed monoculm tillers and highest number of productive tillers were recorded for FBMRS-31 (8) and FBMRS-24 (10) and average 49 lines recorded average number of tillers i.e., 3.49. Green fodder yield was ranged from 7.2 t/ha to 54.4 t/ha. The lowest green fodder was yielded by FBMRS-162 (7.2 t/ha) and highest green fodder yielder was FBMRS-31 (54.4 t/ha) and average green fodder yield was recorded 29.5 tons/ha. The average refractometer brix percentage was 13.5%, however, four lines viz., FBMRS-58, FBMRS-15, FBMRS-173, FBMRS-106 and FBMRS-24 were recoded highest brix percentage (18%), and line FBMRS-164 recorded lowest brix percentage (5%). Length of internode varies from 15 cm to 30 cm, the highest internode length registered for FBMRS-33 (30 cm) and lowest internode length recorded by FBMRS-54 (15 cm). Among the traits studied more than 20 per cent coefficient of variation had recorded by Number of productive tillers and Green fodder yield (t/ha) (Table 2).

The principal component analysis is an extremely useful tool for studying large number of data and is desired to extract most significant data from those data points. PCA is conducted in a sequence of steps, with subjective decisions being made at many of these steps. The number of components extracted is equal to the number of variables being analysed. The first component can be expected to account for a large amount of the total variance. Each succeeding component accounts for progressively smaller amounts of variance. Data were considered in each component with Eigen values more than 1 as per the suggestions given by Brejda et al. (2000), which determines as a minimum 10 % of the variation. Superior Eigen values are considered as best attributes in principal components. PCA has shown the genetic diversity of the larger population or gene pool. Principal component analysis showed that out of eight components derived first three explained most of the total variations present in the gene pool. The first three principal components with Eigen value > 1 contributed about 62.5 % of the total variability among the 194 brown midrib sorghum advanced breeding lines evaluated for different morphological traits. The remaining three components contributed only 37.4 % towards the total morphological diversity among the breeding lines studied. These results are presented in Table 3. The PC 1 contributed maximum variability of 24.7 % followed by PC 2 shows total phenotypic variability of 20.2 % and PC 3 had contributed 17.5 % of the total variation. The important morphological traits in PC 1 were due to variations among the brown midrib advanced sorghum breeding lines but had negative factor loading values for all traits, this indicates all traits contributed negatively. PC 2 was related to diversity among the brown midrib advanced sorghum breeding lines due to Days to 50% flowering, plant height (cm) and Internode length (cm). Similarly, PC

TABLE 1 List of advanced BMR Sorghum breeding lines used in the study

S. No.	Line code								
1	FBMRS-1	48	FBMRS-48	95	FBMRS-95	142	FBMRS-142	189	FBMRS-189
2	FBMRS-2	49	FBMRS-49	96	FBMRS-96	143	FBMRS-143	190	FBMRS-190
3	FBMRS-3	50	FBMRS-50	97	FBMRS-97	144	FBMRS-144	191	FBMRS-191
4	FBMRS-4	51	FBMRS-51	98	FBMRS-98	145	FBMRS-145	192	FBMRS-192
5	FBMRS-5	52	FBMRS-52	99	FBMRS-99	146	FBMRS-146	193	FBMRS-193
6	FBMRS-6	53	FBMRS-53	100	FBMRS-100	147	FBMRS-147	194	FBMRS-194
7	FBMRS-7	54	FBMRS-54	101	FBMRS-101	148	FBMRS-148		
8	FBMRS-8	55	FBMRS-55	102	FBMRS-102	149	FBMRS-149		
9	FBMRS-9	56	FBMRS-56	103	FBMRS-103	150	FBMRS-150		
10	FBMRS-10	57	FBMRS-57	104	FBMRS-104	151	FBMRS-151		
11	FBMRS-11	58	FBMRS-58	105	FBMRS-105	152	FBMRS-152		
12	FBMRS-12	59	FBMRS-59	106	FBMRS-106	153	FBMRS-153		
13	FBMRS-13	60	FBMRS-60	107	FBMRS-107	154	FBMRS-154		
14	FBMRS-14	61	FBMRS-61	108	FBMRS-108	155	FBMRS-155		
15	FBMRS-15	62	FBMRS-62	109	FBMRS-109	156	FBMRS-156		
16	FBMRS-16	63	FBMRS-63	110	FBMRS-110	157	FBMRS-157		
17	FBMRS-17	64	FBMRS-64	111	FBMRS-111	158	FBMRS-158		
18	FBMRS-18	65	FBMRS-65	112	FBMRS-112	159	FBMRS-159		
19	FBMRS-19	66	FBMRS-66	113	FBMRS-113	160	FBMRS-160		
20	FBMRS-20	67	FBMRS-67	114	FBMRS-114	161	FBMRS-161		
21	FBMRS-21	68	FBMRS-68	115	FBMRS-115	162	FBMRS-162		
22	FBMRS-22	69	FBMRS-69	116	FBMRS-116	163	FBMRS-163		
23	FBMRS-23	70	FBMRS-70	117	FBMRS-117	164	FBMRS-164		
24	FBMRS-24	71	FBMRS-71	118	FBMRS-118	165	FBMRS-165		
25	FBMRS-25	72	FBMRS-72	119	FBMRS-119	166	FBMRS-166		
26	FBMRS-26	73	FBMRS-73	120	FBMRS-120	167	FBMRS-167		
27	FBMRS-27	74	FBMRS-74	121	FBMRS-121	168	FBMRS-168		
28	FBMRS-28	75	FBMRS-75	122	FBMRS-122	169	FBMRS-169		
29	FBMRS-29	76	FBMRS-76	123	FBMRS-123	170	FBMRS-170		
30	FBMRS-30	77	FBMRS-77	124	FBMRS-124	171	FBMRS-171		
31	FBMRS-31	78	FBMRS-78	125	FBMRS-125	172	FBMRS-172		
32	FBMRS-32	79	FBMRS-79	126	FBMRS-126	173	FBMRS-173		
33	FBMRS-33	80	FBMRS-80	127	FBMRS-127	174	FBMRS-174		
34	FBMRS-34	81	FBMRS-81	128	FBMRS-128	175	FBMRS-175		
35	FBMRS-35	82	FBMRS-82	129	FBMRS-129	176	FBMRS-176		
36	FBMRS-36	83	FBMRS-83	130	FBMRS-130	177	FBMRS-177		
37	FBMRS-37	84	FBMRS-84	131	FBMRS-131	178	FBMRS-178		
38	FBMRS-38	85	FBMRS-85	132	FBMRS-132	179	FBMRS-179		
39	FBMRS-39	86	FBMRS-86	133	FBMRS-133	180	FBMRS-180		
40	FBMRS-40	87	FBMRS-87	134	FBMRS-134	181	FBMRS-181		
41	FBMRS-41	88	FBMRS-88	135	FBMRS-135	182	FBMRS-182		
42	FBMRS-42	89	FBMRS-89	136	FBMRS-136	183	FBMRS-183		
43	FBMRS-43	90	FBMRS-90	137	FBMRS-137	184	FBMRS-184		
44	FBMRS-44	91	FBMRS-91	138	FBMRS-138	185	FBMRS-185		
45	FBMRS-45	92	FBMRS-92	139	FBMRS-139	186	FBMRS-186		
46	FBMRS-46	93	FBMRS-93	140	FBMRS-140	187	FBMRS-187		
	FBMRS-47	94	FBMRS-94	141		188			

3 expressed positive loading values for variations among the brown midrib advanced sorghum breeding lines resulted from Plant height (cm) and green fodder yield (t/ha).

Scree plot explained the percentage of variation associated with each principal component

obtained by drawing a graph between Eigen values and principal component numbers. PC1 showed 24.7% variability with the Eigen value of 1.48. The Eigen values are gradually declined from PC1 to PC6. The Eigen values for remaining principal components were 1.213, 1.056, 0.942, 0.703 and 0.603 for PC2, PC3,

Variables	Sum	Mean	S.D.	C.V.	Minimum	Maximum
	1.5010					
Days to 50% flowering	15010	77.37	2.96	0.04	74	88
Plant height (cm)	34537	178.03	14.04	0.08	148	225
Number of productive tillers	677	3.49	1.37	0.39	1	10
Internode length (cm)	4307	22.20	2.76	0.12	15	30
Refractometer brix percentage	2623	13.52	2.64	0.20	5	18
Green fodder yield (t/ha)	5733.8	29.56	9.08	0.31	7.2	54.4

TABLE 2
Characteristic means and variations for 194 brown midrib Sorghum advanced breeding lines

PC4, PC5 and PC6 respectively (Fig. 1). The distribution of the brown midrib advanced sorghum breeding lines accounted by different variables from component 1 and component 2 (lines arranged by their line code) were distributed in different groups, which clearly showed genetic diversity among the brown midrib advanced sorghum breeding lines (Fig. 2 to 3). Kassahun TESFAYE, 2017 utilized 117 Sorghum accessions and two standard checks for genetic diversity study. The study found that first four principal components (PCs) with eigenvalues greater than 1 explained about 71.9% of the total variation among accessions for all traits evaluated. The first principal component (PC1) obtained from the study was 26.9% and responsible loading factors were Leaf number at maturity and Plant height. Ahmed et al. (2015) estimated the genetic divergence among 127 sorghum genotypes and reported 83.38% total variation was observed for first six components and remaining two components were responsible for 46.11% variation only.

For considering a minimum threshold Eigen value of one, the first three components accounted for a cumulative of about 62.5% of the whole phenotypic diversity observed among the germplasm accessions. These findings are in accordance with the earlier findings of Nachimuthu *et al.*, 2014; Atul Kumar *et al.*, 2017 and Abraha *et al.*, 2015. Principal

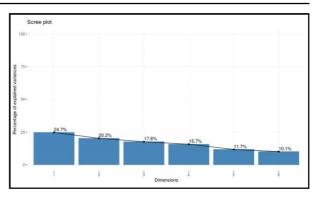


Fig. 1. Scree plot for Eigen values and principal components.

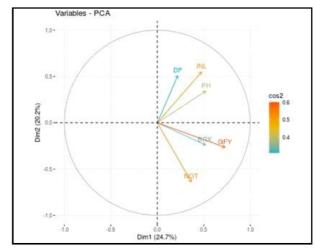


Fig. 2. Variables principal components.

 ${\it TABLE~3} \\ {\it Principal component analysis for different morphological traits recorded in brown midrib Sorghum advanced breeding lines}$

Principal Component	PC1	PC2	PC3	PC4	PC5	PC6
Eigen value	1.483	1.213	1.056	0.942	0.703	0.603
% of total variance	24.719	20.213	17.594	15.704	11.712	10.058
Cumulative variance	24.719	44.932	62.526	78.23	89.942	100
Factor loading by different mor	phological trait	s				
Days to 50% flowering	-0.181	0.462	-0.703	0.064	-0.355	0.359
Plant height (cm)	-0.429	0.31	0.487	0.341	-0.546	-0.261
Number of productive tillers	-0.301	-0.58	-0.315	-0.38	-0.43	-0.381
Internode length (cm)	-0.392	0.498	-0.028	-0.522	0.423	-0.383
Refractometer brix percentage	-0.426	-0.218	-0.313	0.652	0.444	-0.224
Green fodder yield (t/ha)	-0.599	-0.241	0.263	-0.195	0.122	0.679

TABLE 4

Distribution of 194 brown midrib Sorghum advanced breeding lines into twenty clusters

Clusters	Cluster size	Line in cluster
1	2	FBMRS-24, FBMRS-31
2	2	FBMRS-42, FBMRS-43
3	9	FBMRS-64, FBMRS-79, FBMRS-106, FBMRS-107, FBMRS-124, FBMRS-140, FBMRS-142, FBMRS-154, FBMRS-170
4	3	FBMRS-87, FBMRS-88, FBMRS-167
5	17	FBMRS-19, FBMRS-20, FBMRS-51, FBMRS-56, FBMRS-58, FBMRS-59, FBMRS-74, FBMRS-75, FBMRS-76, FBMRS-134, FBMRS-135, FBMRS-159, FBMRS-168, FBMRS-169, FBMRS-177, FBMRS-184, FBMRS-187
6	7	FBMRS-100, FBMRS-141, FBMRS-144, FBMRS-145, FBMRS-147, FBMRS-158, FBMRS-191
7	13	FBMRS-1, FBMRS-34, FBMRS-35, FBMRS-50, FBMRS-62, FBMRS-98, FBMRS-105, FBMRS-125, FBMRS-128, FBMRS-130, FBMRS-143, FBMRS-180, FBMRS-186
8	10	FBMRS-16, FBMRS-23, FBMRS-26, FBMRS-28, FBMRS-41, FBMRS-46, FBMRS-53, FBMRS-67, FBMRS-83, FBMRS-84
9	9	FBMRS-109, FBMRS-118, FBMRS-126, FBMRS-127, FBMRS-155, FBMRS-156, FBMRS-171, FBMRS-178, FBMRS-183
10	7	FBMRS-92, FBMRS-160, FBMRS-161, FBMRS-162, FBMRS-163, FBMRS-164, FBMRS-165
11	10	FBMRS-6, FBMRS-8, FBMRS-30, FBMRS-37, FBMRS-38, FBMRS-45, FBMRS-65, FBMRS-86, FBMRS-103, FBMRS-112
12	9	FBMRS-13, FBMRS-25, FBMRS-114, FBMRS-119, FBMRS-123, FBMRS-138, FBMRS-151, FBMRS-174, FBMRS-194
13	11	FBMRS-18, FBMRS-29, FBMRS-36, FBMRS-48, FBMRS-49, FBMRS-55, FBMRS-57, FBMRS-85, FBMRS-173, FBMRS-175, FBMRS-182
14	8	FBMRS-2, FBMRS-3, FBMRS-7, FBMRS-11, FBMRS-94, FBMRS-132, FBMRS-185, FBMRS-193
15	11	FBMRS-44, FBMRS-54, FBMRS-61, FBMRS-63, FBMRS-80, FBMRS-81, FBMRS-90, FBMRS-91, FBMRS-99, FBMRS-166, FBMRS-176
16	13	FBMRS-9, FBMRS-14, FBMRS-15, FBMRS-17, FBMRS-39, FBMRS-47, FBMRS-60, FBMRS-68, FBMRS-70, FBMRS-71, FBMRS-102, FBMRS-136, FBMRS-192
17	11	FBMRS-10, FBMRS-21, FBMRS-22, FBMRS-27, FBMRS-52, FBMRS-66, FBMRS-73, FBMRS-93, FBMRS-97, FBMRS-137, FBMRS-149
18	13	FBMRS-4, FBMRS-69, FBMRS-89, FBMRS-101, FBMRS-108, FBMRS-110, FBMRS-113, FBMRS-116, FBMRS-146, FBMRS-148, FBMRS-150, FBMRS-152, FBMRS-157
19	17	FBMRS-5, FBMRS-12, FBMRS-32, FBMRS-33, FBMRS-40, FBMRS-77, FBMRS-78, FBMRS-82, FBMRS-95, FBMRS-96, FBMRS-104, FBMRS-111, FBMRS-115, FBMRS-122, FBMRS-133, FBMRS-172, FBMRS-189
20	12	FBMRS-172, FBMRS-107 FBMRS-72, FBMRS-117, FBMRS-120, FBMRS-121, FBMRS-129, FBMRS-131, FBMRS-139, FBMRS-153, FBMRS-179, FBMRS-181, FBMRS-188, FBMRS-190

component 1 explained maximum variability from total variation for morphological traits and PC 2 loaded with 20.2% of total phenotypic variability and PC 3 had contributed 17.5% of the total variation. By using Hierarchical Clustering method 194 brown midrib advanced sorghum breeding lines grouped into twenty distinct clusters (Table 4). The minimum distance between the clusters various between 0.5 to 8.0 shown in Fig. 4, Among the clusters, cluster 5 and 19 had the highest lines (17 each cluster) followed by cluster 7, 17 and 18 had 13 lines each and cluster 20 holded 12 lines. But cluster 1, 2 and 4 had only 2-3 lines. Grouping of lines into different clusters confirmed the presence of variation among genotypes. Maji and Shaibu, 2012 reported that germplasm evaluation and characterization

is a routine endeavour for plant breeders, and application of PCA tool, cluster and multivariate statistical analysis provide a useful means for estimating morphological diversity within and between germplasm collections. These tools are useful for the evaluation of potential breeding value and used to detect significant differences between germplasm and magnitude of deviation among crop species.

Results of the study revealed that there is a large quantity of variability present in brown midrib advanced sorghum breeding lines. PCA identified only few characters played important role in classifying the variation present in these BMR sorghum lines. The results of the PCA revealed that the 62.5% of the total variability was explained by the first three principal

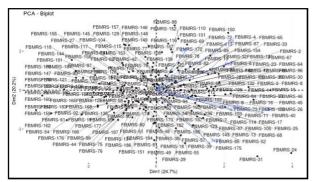


Fig. 3. Distribution of genotypes across first two components based on PCA.

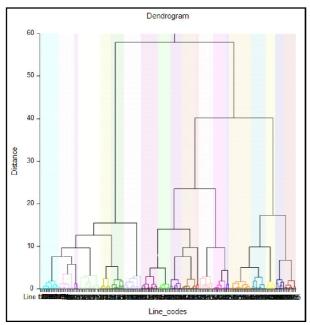


Fig. 4. Phenotypic dendrogram generated by Hierarchical Clustering method.

components. The same trend was observed in grouping of genotypes by clustering analysis. Combination of both PCA and clustering provides the absolute results of variability existed in the gene pool (Fig.4). Among the twenty clusters, Cluster number 1 (FBMRS-24 and FBMRS-31) had only two lines each and both were to be superior for highest green fodder yield (tons/ha) (FBMRS-24 (48.7 t/ha) and FBMRS-31 (54.4 t/ha) and refractometer brix percentage 18% and 17%, respectively. Selection of parents based on these findings will produce more of genetic diversity in the crop evolution. Highest level of variability existing in the genotypes and important traits will open the scope for additional enhancement of the cultivars in crop improvement programmes in brown midrib sorghum breeding program.

Principal component analysis was utilized to examine the variation and to estimate the relative

contribution of various traits for total variability. The present study identified seven best performing brown midrib sorghum breeding lines viz., FBMRS-24, FBMRS-31, FBMRS-42,

FBMRS-43, FBMRS-87, FBMRS-88 and FBMRS-167 for considering the green fodder yield coupled with higher brix percentage. These brown midrib sorghum breeding lines may be utilized in the recombination breeding programmes to develop high yielding new brown midrib sorghum varieties and hybrid as well.

ACKNOWLEDGEMENT

The author would like to thank the Director, Foragen Seeds Private Limited, Hyderabad, Telangana, India, Mr. Anil Reddy and Mr. Prabhakar babu Gunturu, CEO, Foragen Seeds Private Limited, Hyderabad for their great support of funding and special thanks to Mrs. Radha Madhvi, HR, Foragen Seeds Private Limited for her administrative support for this research investigation.

REFERENCES

Ahamed, K. U., B. Akhter, M.R. Islam, M.K. Alam and M.M. Hossain, 2015: An assessment of genetic diversity in sorghum (*Sorghum bicolor L. Moench*) germplasm. *Bull. Inst. Trop. Agr.*, Kyushu Univ., **38**: 47-54.

Brejda, J.J., T.B. Moorman, D.L. Karlen. and T.H. Dao, 2000: Identification of regional soil quality factors and indicators. I. Central and Southern High-Plains. *Soil Sci. Soc. Am. J.*, **64**: 2115-2124.

Gopinath, P. P., R. Parsad, B. Joseph, and V. S. Adarsh, 2020: GRAPES: General R-shiny Based Analysis Platform Empowered by Statistics. kaugrapes. com, version 1.0.0, DOI: 10.5281/zenodo.4923220

Jolliffe, I.T. and J. Cadima, 2016: Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A.*, **374**: 20150202.

Kassahun, TESFAYE., 2017: Genetic diversity study of sorghum (Sorghum bicolor (L.) Moench) genotypes, Ethiopia. Acta Universitatis Sapientiae Agriculture and Environment, 9: 44-54.

Maji, A.T. and A. A. Shaibu, 2012: Application of principal component analysis for rice germplasm characterization and evaluation. *J. Plant Breed. Crop Sci.*, **4**: 87-93.

Nachimuthu, V. V., S. Robin, D. Sudhakar, M. Raveendran, S. Rajeswari and S. Manonmani, 2014: Evaluation of rice genetic diversity and variability in a population panel by principal component analysis. *Indian J. Sci. Technol.*, **10**: 1555- 1562.