



P-ISSN: 2349-8528

E-ISSN: 2321-4902

[www.chemijournal.com](http://www.chemijournal.com)

IJCS 2020; 8(5): 2206-2216

© 2020 IJCS

Received: 26-07-2020

Accepted: 29-08-2020

**Ashwini Karadi**Indian Institute of Horticultural  
Research, Bangalore, Karnataka,  
India**ST Kajjidoni**University of Agricultural  
Science, Dhanwad, Karnataka,  
India

## Principal component analysis for productivity and grain quality traits in *rabi* sorghum (*Sorghum bicolor* L. Moench)

**Ashwini Karadi and ST Kajjidoni**DOI: <https://doi.org/10.22271/chemi.2020.v8.i5ad.10636>

### Abstract

Principal component analysis was carried out involving 24 selected mini core and 16 promising varieties during *rabi* sorghum. Principal component analysis revealed that, four out of twelve principal components with eigen values >1 were extracted. Among the selected mini core, three components contributed 66.37% of the total variation. Principal components first three contributed 30.55%, 19.74%, and 16.07% toward the variation, respectively. Plant height and true density traits contributed positively in mini core accessions. Whereas, among promising released varieties the three components contributed 66.46% of the total variation and first three contributed, 30.07%, 19.51% and 16.88%, respectively. Ear head width, 100 seed weight, seed volume, true density and seed size has showed positive contribution in all three principal components of variation observed in varieties. So, these characters spread over different principal component and contributing most to explain the variability should be taken into consideration for their utilization in breeding programme.

**Keywords:** Sorghum, mini core, principal component analysis, multivariate, eigen value

### Introduction

Sorghum (*Sorghum bicolor* L. Moench) is a tropical cereal crop grown in a wide range of environments, where it is one of the staple food of many people. It is the fifth most important cereal crop after wheat (*Triticum aestivum*), maize (*Zea mays*), rice (*Oryza sativa*), and barley (*Hordeum vulgare*). Worldwide, it is cultivated on 41.07 million ha area with production of 58.42 million tonnes in the year approx., 2019-20 (Anonymous, 2019a) <sup>[1]</sup>. In India, sorghum having 5.00 million ha area with 4.5 million tonnes production and 900 kg/ha productivity in the year 2019-20 (Anonymous, 2019b) <sup>[2]</sup>. *Rabi* sorghum is highly valued as a food crop because of its excellent grain quality, nutritional status and low glycemic index and gluten-free (Taylor *et al.*, 2006) <sup>[3]</sup>. It is grown mainly for human consumption about more than 35% of its production; the rest is used primarily for animal feed, alcohol production and industrial products (FAO, 1995; Awika and Rooney, 2004; Dicko *et al.*, 2006; Mehmood *et al.*, 2008) <sup>[4, 5, 6, 7]</sup>. As a food crop, sorghum can provide a healthy diet for human nutrition (Mudjisihi and Damardjati, 1987) <sup>[8]</sup>. Nutritional qualities like seed protein and seed amylose content are the important aspects, where there is necessity and scope to improve nutritional quality along with the grain quality traits of *rabi* sorghum. Multivariate analysis such as principal component analysis (PCA) and cluster analysis serve as potential tools in evaluating the phenotypic diversity, identifying genetically distant clusters of genotypes and selecting important traits contributing to the total variation in the germplasm. Principal component analysis (PCA) and two-way cluster analysis are two important statistical programs that aid in selecting elite germplasm lines for breeding program that meet the goal of a breeder for the development of improved varieties (Mohammadi and Prasanna, 2003) <sup>[9]</sup>.

Principal component analysis aim is to resolve the total variation of a set of traits into linear, independent composite traits, which successively maximize variability in the data (Johnson, 2012) <sup>[10]</sup>. It is the most common technique used in variability studies and numerical classification; to discriminate crop genotypes (Das, 2000) <sup>[11]</sup> and (Yan and Kang, 2003) <sup>[12]</sup>. It allows natural grouping of the genotypes and is precise indicator of differences among genotypes (Bello, 2004) <sup>[13]</sup>.

**Corresponding Author:****Ashwini Karadi**Indian Institute of Horticultural  
Research, Bangalore, Karnataka,  
India

The PCA or canonical root analysis is technique attempt to simplify and analyze the inter relationship among a large set of variables in term of a relatively a small set of variables or components without losing any essential information of original data set. It may be used to disclose the patterns and eradicate redundancy in data sets as variations regularly arise in crop species for yield (Maji and Shaibu, 2012) <sup>[14]</sup> and (Adams, 1995) <sup>[15]</sup>. Since, PCA extract all the important components and highlight their contribution toward the total variability, it can be the choice as an important tool to speed up the breeding programme.

### Materials and Methods

The plant material for this experiment comprised of (a) 24 selected mini core collection out of 208 mini core accessions based on grain hardness, grain colour and grain lustre along

with M 35-1 as check variety (Table-1). The mini core collection of sorghum obtained from DSR Hyderabad. (b) Sixteen selected released/promising varieties of *rabi* sorghum which are commonly grown in northern Karnataka (Table-2). The experiment was laid out in medium deep black soil under rain fed condition was carried out in two experiments at AICRP, UAS, Dharwad during *rabi* season of 2012-13. The randomized block design was followed separately with two replications and each entry was sown in four rows of 4 m length with inter row spacing of 45 cm and intra row spacing of 15 cm. Observations on all quantitative characters like plant height (cm), panicle length (cm), panicle width (cm), seed yield per plot (g), 100 seed weight (g), seed volume (ml), bulk density (g/ml), true density (g/ml), seed size (mm), seed protein (%), seed amylose (%), seed amylopectin (%) and seed yield/plot (kg).

**Table 1:** Selected genotypes of mini core collections of sorghum (Experiment-a)

S. No.	Accession number	Origin	Seed hardness	Seed lustre	Seed color
1	IS-473	USA	Very hard	Lustrous	White
2	IS-1041	India	Hard	Lustrous	White
3	IS-2379	South Africa	Very hard	Non lustrous	Light brown
4	IS-3971	India	Very hard	Intermediate	Creamy straw
5	IS-4515	India	Hard	Lustrous	White
6	IS-5295	India	Very hard	Intermediate	Chalky white
7	IS-5301	India	Very hard	Lustrous	White
8	IS-12697	Australia	Hard	Intermediate	Brown
9	IS-12937	Ethiopia	Very hard	Lustrous	Light red
10	IS-12945	Nicaragua	Very hard	Lustrous	White
11	IS-13294	Venezuela	Very hard	Intermediate	Light brown
12	IS-13459	Mexico	Hard	Lustrous	Brown
13	IS-13971	South Africa	Very hard	Intermediate	Light brown
14	IS-15931	Cameroon	Hard	Lustrous	Chalky white
15	IS-19153	Sudan	Very hard	Lustrous	White
16	IS-22720	Somalia	Hard	Lustrous	White
17	IS-24139	Tanzania	Very hard	Intermediate	White
18	IS-28849	Yemen	Hard	Intermediate	White
19	IS-29358	Lesotho	Very hard	Intermediate	Light yellow
20	IS-30443	China	Very hard	Lustrous	Chalky white
21	IS-30450	China	Very hard	Non lustrous	Brown
22	IS-30572	Cameroon	Very hard	Intermediate	Yellow
23	IS-13893	South Africa	Hard	Intermediate	Reddish brown
24	IS-13782	South Africa	Very hard	Intermediate	Red
25	M35-1	India	Hard	Lustrous	White

**Table 2:** List of selected released/promising varieties of *rabi* sorghum (Experiment-b)

S. No.	Varieties name
1	DSV-4
2	SPV-86
3	SPV-1829
4	BJV-44
5	DSV-5
6	A-1
7	CSV-216R(Phuleyashoda)
8	PhuleVasudha
9	PhuleRevathi
10	M35-1 (Akola source)
11	M35-1 (Bijapur source)
12	BarsiJowar
13	Kodmurki (popular local)
14	SVD-803(Advanced breeding lines)
15	SVD-808 (Advanced breeding lines)
16	SVD-770(Advanced breeding lines)

### Grain quality parameters

Seed size was measured by using *Vernier Callipers* where length, breadth and thickness of seeds were recorded. Seed

density classified into two types *viz.*, true density and bulk density. Seed bulk density was measured by hundred gram of seeds were weighed and volume was recorded in a measuring jars. Whereas, seed true density was observed by known weight of seeds placed in a measuring jar containing known quantity of toluene. Increase in volume was recorded after pouring seeds in measuring jar. Seed volume was noted with countable numbers of seeds were placed in a measuring jar. Grain quality characters like seed luster, seed color, seed shape and seed hardness was recorded by measuring the grinding time required to obtain a fixed volume of flour from the grains.

### Estimation of biochemical parameters

**(a) Seed protein (per cent):** Protein content of selected genotypes was estimated by using Microkjeldhal method. Total nitrogen was estimated by using Kel-plus (digestion and distillation unit). Crude protein value was obtained by multiplying the total nitrogen by the conversion factor.

### (b) Amylose content and amylopectin content (percent):

Total amylose was estimated by following the method of Soubhagya and Bhattacharya (1979) <sup>[16]</sup>, 100 mg sample was

taken in a 100 ml volumetric flask, disperse 1 ml of alcohol followed by 10 ml of 1 N NaOH leave it for overnight, make the volume upto 100 ml from distilled water. From this extract 2.5 ml was taken, add 20ml distilled water, add 3 drops of phenolphthalein indicator, it will change to pink color, add 0.1 N HCl, till it becomes colourless, now add 1 ml of 0.2 per cent iodine solution, make volume made up to 100 ml. The purple-blue was read at 590 nm. Amylopectin will be calculated by subtracting total amylose from 100.

**Principal Component Analysis:** In order to study the proportion of variance explained by each yield characteristics Principal components analysis was used. This technique was first described by Karl Pearson (1901) [17] and the description of practical computing methods came much later from Hotelling (1933) [18]. It involves a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components (Chatfield and Collis 1980) [19]. Principal component analysis is a multivariate technique for examining the relationships among several quantitative variables.

The object of the analysis is to take  $p$  variables  $X_1, X_2, \dots, X_p$  and find combinations of these to produce indices  $Z_1, Z_2, \dots, Z_p$  that are uncorrelated. The absence of correlation means that the indices are measuring from different dimensions in the data. The indices are ordered so that  $Z_1$  displays the large amount of variation,  $Z_2$  displays the second largest amount of variation and so on.

That is,  $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \text{var}(Z_3) \geq \dots \geq \text{var}(Z_p)$ , where  $\text{var}(Z_i)$  denotes the variance of  $Z_i$  in the data set. The  $Z_i$  are called principal components. Principal component analysis depends only on the covariance matrix  $\Sigma$  or the correlation matrix of the variable under study. The best results are obtained when the original variables are very highly correlated, positively or negatively. In principal component analysis, the variance of most of the indices is low as to be negligible. In that case the variation in the data set can be described by the few  $Z$  variables with variances that are not negligible.

Let  $X_1, X_2, \dots, X_p$  are variables under study, then the first principal component is the linear combination of the variables  $X_1, X_2, \dots, X_p$ ,

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p^2 = 1$$

Such that variance of  $Z_1$  is as large as possible subject to the condition that

$$a_{11}^2 + a_{12}^2 + \dots + a_{1p}^2 = 1$$

This constraint is introduced because if this is not done, then  $\text{var}(Z_i)$  can be increased simply by increasing any one of the  $a_{ij}$  values. The second principal component,

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p$$

is such that variance of  $Z_2$  is as large as possible next to  $\text{var}(Z_1)$  subject to the condition that

$$a_{21}^2 + a_{22}^2 + \dots + a_{2p}^2 = 1 \text{ and } \text{Cov}(Z_1, Z_2) = 0$$

The third principal component,

$$Z_3 = a_{31}X_1 + a_{32}X_2 + \dots + a_{3p}X_p$$

is such that variance of  $Z_3$  is as large as possible next to  $\text{var}(Z_1)$  and  $\text{var}(Z_2)$ .

PCA analysis of yield and quality characteristics of sorghum analyzed statistically using the software WINDOSTAT, developed by INDOSTAT services Ltd. Hyderabad, India.

## Results and Discussion

The objectives of PCA are to discover or to reduce the dimensionality of the data set and to identify new meaningful underlying variables (Jolliffe, 2002) [20]. Watson and Eyzaguirre, (2002) [21] also reported that PCA of morphological characterization results could identify a few key or minimum descriptors that effectively account for the majority of the diversity observed, saving time and effort for future characterization efforts. Principal components approach is very helpful in deciding which agronomic traits of crop contributing most to yield, subsequently, these agronomic traits should be emphasized in the breeding program (Jain *et al.*, 2016) [22]. The PCA reduce relatively a large series of data into smaller number of components by looking for groups that have very strong inter-correlation in a set of variables and each component explained per cent (%) variation to the total variability. Large data sets are increasingly common and are often difficult to interpret. Principal component analysis is an analytical method for transforming multiple indexes into fewer new ones. It does so by creating new uncorrelated variables that successively maximize variance. 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, hence making PCA an adaptive data analysis technique (Jolliffe and Cadima, 2016) [23]. PCA can be used to find out the resemblance between the variables and classify the genotypes (Leonard and Peter, 2009) [24].

Using the data on plant biometrical characters  $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}$  and  $X_{12}$  the principal component analysis has been carried out for the data of both the experiments. Four out of twelve principal components with eigen values  $> 1$  were extracted. Among selected mini core collection, these three components contributed 66.37% of the total variation among the mini core, also from Figure 2 it is evident graphically that first principal component extracted maximum amount of variation among yield characteristics in the data. Principal components 1, 2, and 3 contributed 30.55%, 19.74%, and 16.07% toward the variation observed among accessions, respectively (Table 3; Fig 1 and 2). Whereas, among selected promising released varieties the three components contributed 66.46% of the total variation among the varieties and principle components first three contributed, 30.07%, 19.51% and 16.88%, respectively toward the variation observed among varieties (Table 5 and Fig 6). The Eigenvectors decreased significantly from principal component 2 from 19.74% to 16.07% in mini core and 19.51% to 16.88%. This suggests that after principal component 3 more principal components did not describe much variation. The first principal component is the largest contributor to the total variation in the population followed by subsequent components. The criteria used by Clifford and Stephenson (1975) [25] and corroborated by Guei *et al.* (2005) [26], suggested that the first three principal components are often the most important in reflecting the variation patterns among accessions, and the characters associated with these are more useful in differentiating the accessions. Thus it is useful for genetic improvement of important traits having larger contributions to the variability rather than going for all

the characters under study. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. Hence, only the first three PCs were considered. When compared to both the set the principal components contributed nearer values, it might be result of conscious selection towards higher yield in breeding program.

An excellent way to see the results of PCA are through graphics Biplot (Gabriel 1971) [27], specifically the connections between the ordinations of the rows (generally samples) and of the columns (generally markers) of the data table. The Biplot represent both ordinations simultaneously. The mathematical technique used in PCA is called eigen analysis: to solve for the eigen values and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigen value has the same direction as the first principal component. The eigenvector associated with the second largest eigen value determines the direction of the second principal component. The sum of the eigen values equals the trace of the square matrix and the maximum number of eigenvectors equals the number of rows (or columns) of this matrix (Harris 2001) [28]. Multivariate statistical methods have been successfully used to classify quantitative and qualitative variations in many crop species like pea reported by Amurrio *et al.* (1995) [29]; Rabbani *et al.* (1998) [30]; Berdahl *et al.* (1999) [31]; Chandran and Padya (2000) [32]; Cravero *et al.* (2002) [33]; Rosso and Pagano (2005) [34] and Bhargava *et al.* (2007) [35].

Among selected mini core collection, variation relative to the first component was associated with plant height (0.1257), ear head length (0.3227), ear head width (0.2846), true density (0.1217), seed amylose content (0.2235). The second principal component variable contributing most positively were associated with plant height (0.0089), seed yield per plot (0.1197), 100 seed weight (0.2494), seed volume (0.1033), seed bulk density (0.4757), seed true density (0.2734), seed size (0.0391) and amylose (0.4232). Whereas, the third principle component was associated with all traits except seed protein content (-0.1318) and seed amylose content (-0.344). The fourth principal component variable contributing most positively was associated with bulk density (0.1735), true density (0.3605) and amylopectin (0.344). Among selected promising released varieties, variation relative to the first component was contributing most positively with all traits except plant height (-0.2753), ear head length (-0.2467), seed protein content (-0.1931) and seed amylopectin (-0.4351) were negatively loaded. The second principal component was contributing most positively with all traits except seed yield per plot (-0.0762), seed bulk density (-0.1057) and seed amylopectin (-0.1538) were negatively loaded. The third principle component was associated positively with all traits except plant height (-0.1821), seed protein content (-0.2718) and seed amylose content (-0.3278) were negatively loaded. Whereas, plant height (0.1932), ear head length (0.4703), ear head width (0.0352), seed yield /plot (0.3140), seed volume (0.5510), protein (0.0385), and amylose (0.0472) contributing positively. The PCA scores for 24 selected mini core along with local check and 16 varieties of sorghum in the first four principal components were computed and were considered as PC1, PC2, PC3 and PC4 squared distance of each genotype from these three axes were calculated (Table 4 and 5). These scores can be utilized to propose precise selection indices whose intensity can be decided by variability explained by

each of principal component. High PC score for a particular genotype in a particular component denotes high values for the variables in that particular genotype.

A scree plot is a simple line segment plot that shows the fraction of total variance in the data. It is a plot, in descending order of magnitude, of the eigen values of a correlation matrix. According to Chatfield and Collins (1980) [19], components with an eigen value of <1 should be eliminated so that fewer components are dealt with. Sharma (1998) [36] reported that PCA reflects the importance of the largest contributor to the total variation at each axis of differentiation. It was further reported by Fenty (2004) [37] that PCA reduces a large set of variables to come up with smaller sets of components those summaries the correlations. The Scree plot of the PCA shows that the first four eigen values correspond to the whole percentage of the variance in the dataset (Fig 3 and 7). The distribution of varieties based on the first and second principal component exhibited the phenotypic variation among the population and explains how these widely dispersed along both the axes. All the accessions and varieties were widely scattered across different quarters (Fig 4 and 8). In this study, we choose to follow the criterion used by Clifford and Stephenson (1975) [25] and result was corroborated with findings of Guei *et al.*, (2005) [26], which suggested that the first three principal components are often the most important in reflecting the variation patterns among varieties, and the characters associated with these are more useful in differentiating characters.

Among selected mini core collection, plant height and true density contributed the highest in variation. Whereas, ear head width, 100 seed weight, seed volume, true density and seed size contributed the highest in variation in selected promising released varieties. Thus, the predominant character coming together in different principal components and contributing towards explaining the variability and have the tendency to remain together. This may be taken into consideration during utilization of these characters in breeding programme. Significance of Eigen value determines the direction of the new feature space and eigen value determines the magnitude of the PCA components. The higher the eigen value is, the higher will be the variance along a co-variance matrix in the eigen vector direction (Principle component). The phenotypic value of each trait measured the importance and contribution of each component to the total variation, whereas each coefficient of proper vectors indicates the degree of contribution of every original variables with which each principal component is associated. The first 3 components accounted for 66.37% of the total variation among the minicore and 66.46% of the total variation among the varieties. Characters with high variability are expected to provide high level of gene transfer during breeding programmes (Gana, 2006 and Gana, 2013, 2014) [38, 39]. Sample traits are generally inter-correlated to varying degrees and hence not all principal components are needed to summarize the data adequately. In this study, the first three principal components represented a sizeable amount of diversity among the genotypes investigated. This implied that several traits were involved in explaining the variation among the genotypes. Ayana and Bekele (1999) [40] reported significance of first five PCs in the total variability of different agro-morphological traits in sorghum.

Several studies on principal component analysis of different agro-morphological traits in sorghum have been documented. On the other hand, head width, head weight, grain yield per plant, and fresh and dry shoot weight were found to be the



most important traits for drought tolerance in grain sorghum (Ali *et al.*, 2011) <sup>[41]</sup>. Abraham *et al.*, (2015) <sup>[42]</sup> reported four principal components with eigen values greater than one, which explained > 75% of the total variation for grain yield, biomass, stay-green, leaf area, peduncle exertion, days to flowering, and maturity. Chikuta *et al.*, (2015) <sup>[43]</sup> was reported, around 44%, 17%, and 15% variation attributed to first, second, and third principal components, respectively. Karadi and Kajjdoni (2019) <sup>[44]</sup> reported that, three components contributed 58.29% of the total variation among the mini core. Principal components first three contributed, 22.73%, 17.99%, and 15.50%, respectively toward the variation observed among accessions. Variation relative to the first component was associated with seed yield per plant, 100 seed weight, seed volume, bulk density, seed size. The second principal component was associated with plant height, ear head length, ear head width, seed yield per plant, 100 seed weight, seed volume and seed size. The third principle component was associated with ear head width, 100 seed weight, seed yield per plant and seed size.

Among selected released varieties, ear head width, 100 seed weight, seed volume, true density and seed size has showed positive contribution in three principal components of variation. Plant height and true density contributed positively in mini core accessions. Seed true density contributed positively in both the sets of sorghum. Seed true density is related to physical seed dimensions, sphericity and seed shape aspects. Improvement towards grain quality aspects is possible only when there is genetic variability exists for these traits in the crop, which is the basic prerequisite for crop improvement. Thus the principal component analysis was helpful, in revealing the high level of genetic variation existing in the population and explains, which characters contribute for genetic diversity among the genotypes in the population. So, these characters spread over different principal component and contributing most to explain the variability should be taken into consideration for their utilization in breeding programme. It provide information that could help in better selection of parental genotypes with specific traits and in devising breeding strategies for trait improvement.

**Table 3:** Principal component analysis of measured traits in selected mini core accessions and promising varieties of rabisorghum

Selected mini core collection												
Parameters	Plant height (cm)	Panicle length (cm)	Panicle width (cm)	100 seed weight (g)	Seed yield/plot (g)	Seed volume (ml)	Bulk density (g/ml)	True density (g/ml)	Seed size (mm)	Protein (%)	Amylose (%)	Amylopectin (%)
eigenvalue	3.67	2.37	1.93	1.47	0.9	0.59	0.49	0.32	0.16	0.08	0.01	6.55
variance.percent	30.55	19.74	16.07	12.23	7.5	4.95	4.1	2.66	1.36	0.71	0.11	5.46
cumulative.variance.percent	30.55	50.3	66.37	78.61	86.11	91.06	95.16	97.82	99.18	99.89	100	100
Promising released varieties												
Parameters	Plant height (cm)	Panicle length(cm)	Panicle width (cm)	100 seed weight (g)	Seed yield/plot (g)	Seed volume (ml)	Bulk density (g/ml)	True density (g/ml)	Seed size (mm)	Protein (%)	Amylose (%)	Amylopectin (%)
eigenvalue	3.61	2.34	2.03	1.38	0.98	0.64	0.54	0.24	0.17	0.06	0.01	3.3
variance. percent	30.07	19.51	16.88	11.53	8.17	5.34	4.53	1.99	1.39	0.51	0.07	2.75
cumulative.variance.percent	30.07	49.58	66.46	77.99	86.16	91.5	96.03	98.02	99.42	99.93	100	100

**Table 4:** Factor loadings of the study traits of the first three principal components (PCs)

Traits	Selected mini core				Promising released varieties			
	PC1	PC2	PC3	PC4	PC1	PC2	PC3	PC4
Plant height (cm)	0.1257	0.0089	0.3415	-0.5040	-0.2753	0.3582	-0.1821	0.1932
Ear head length (cm)	0.3227	-0.0671	0.3521	-0.3738	-0.2467	0.2354	0.3073	0.4703
Ear head width (cm)	0.2846	-0.1921	0.3892	-0.1499	0.0281	0.533	0.208	0.0352
Seed yield /plot (g)	-0.4348	0.1197	0.0956	-0.1372	0.374	-0.0762	0.2234	0.3140
100 seed weight (g)	-0.2953	0.2494	0.2005	-0.2846	0.1969	0.2307	0.2365	-0.5256
Seed volume (ml)	-0.4643	0.1033	0.1162	-0.1128	0.3483	0.0102	0.1191	0.5510
Bulk density (g/ml)	-0.0147	0.4757	0.2548	0.1735	0.291	-0.1057	0.2242	-0.0705
True density (g/ml)	0.1217	0.2734	0.4714	0.3605	0.249	0.4309	0.0537	-0.2027
Seed size (mm)	-0.4248	0.0391	0.0631	-0.2176	0.0039	0.1168	0.606	-0.0990
Protein (%)	-0.109	-0.4588	-0.1318	-0.2874	-0.1931	0.4639	-0.2718	0.0385
Amylose (%)	0.2235	0.4232	-0.344	-0.3010	0.4351	0.1538	-0.3278	0.0472
Amylopectin (%)	-0.2235	-0.4232	0.344	0.3010	-0.4351	-0.1538	0.3278	-0.0472

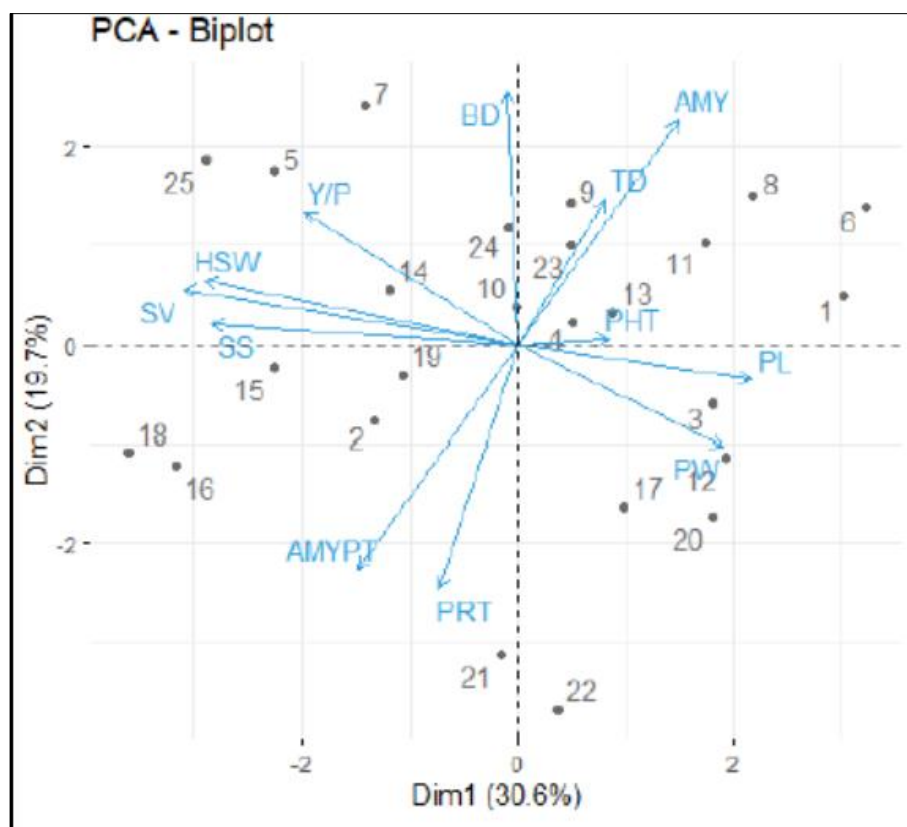
**Table 5:** PCA scores for 24 selected mini core accessions along with check M35-1.

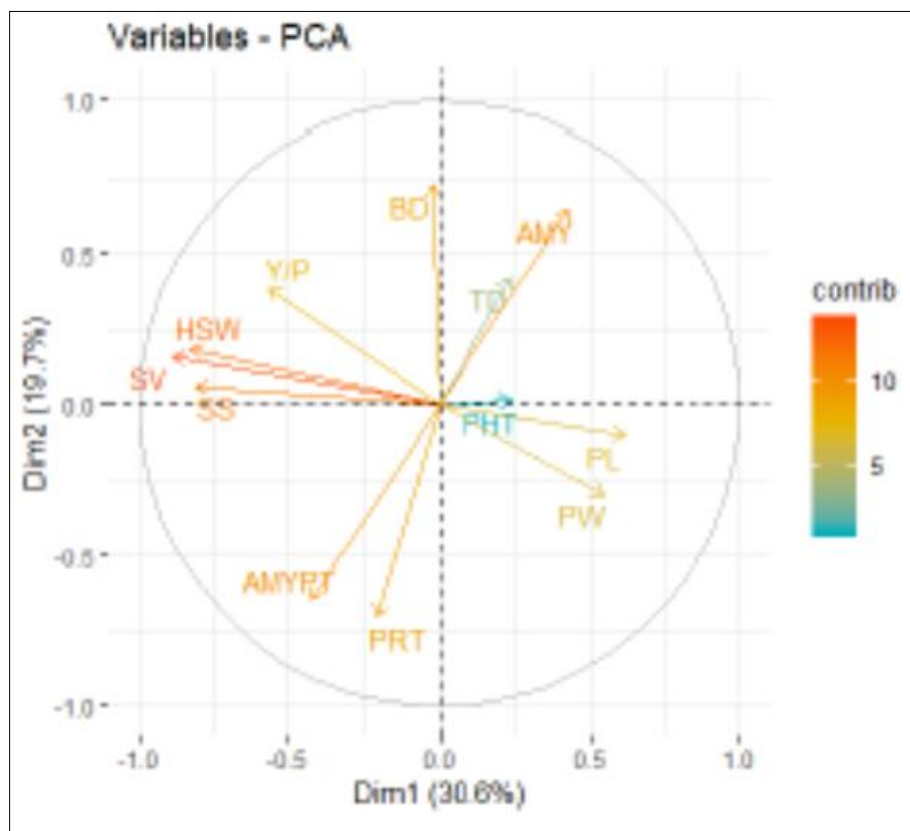
Accession	PC I	PC II	PC III	PC IV
IS-473	3.012	0.499	1.880	0.933
IS-1041	-1.326	-0.775	-2.043	-1.558
IS-2379	1.809	-0.592	-0.292	0.062
IS-3971	0.517	0.225	-2.368	1.462
IS-4515	-2.253	1.742	-0.593	-1.799
IS-5295	3.217	1.392	-0.182	1.090
IS-5301	-1.410	2.416	-0.335	-0.239
IS-12697	2.181	1.497	0.572	1.455
IS-12937	0.492	1.418	0.006	0.553
IS-12945	-0.009	0.377	-0.498	0.177

IS-13294	1.732	1.033	-0.803	-0.464
IS-13459	1.928	-1.140	-1.550	-0.735
IS-13971	0.886	0.319	-1.017	-0.977
IS-15931	-1.190	0.546	1.636	-3.127
IS-19153	-2.263	-0.234	-0.398	0.643
IS-22720	-3.161	-1.213	-0.478	1.048
IS-24139	0.984	-1.626	2.326	-0.981
IS-28849	-3.610	-1.087	0.815	1.362
IS-29358	-1.070	-0.317	0.802	1.958
IS-30443	1.803	-1.724	2.506	-1.007
IS-30450	-0.161	-3.134	-1.599	0.678
IS-30572	0.362	-3.674	0.512	-0.216
IS-13893	0.500	1.002	-0.251	-0.586
IS-13782	-0.076	1.179	-1.308	-0.674
M35-1	-2.893	1.872	2.662	0.943

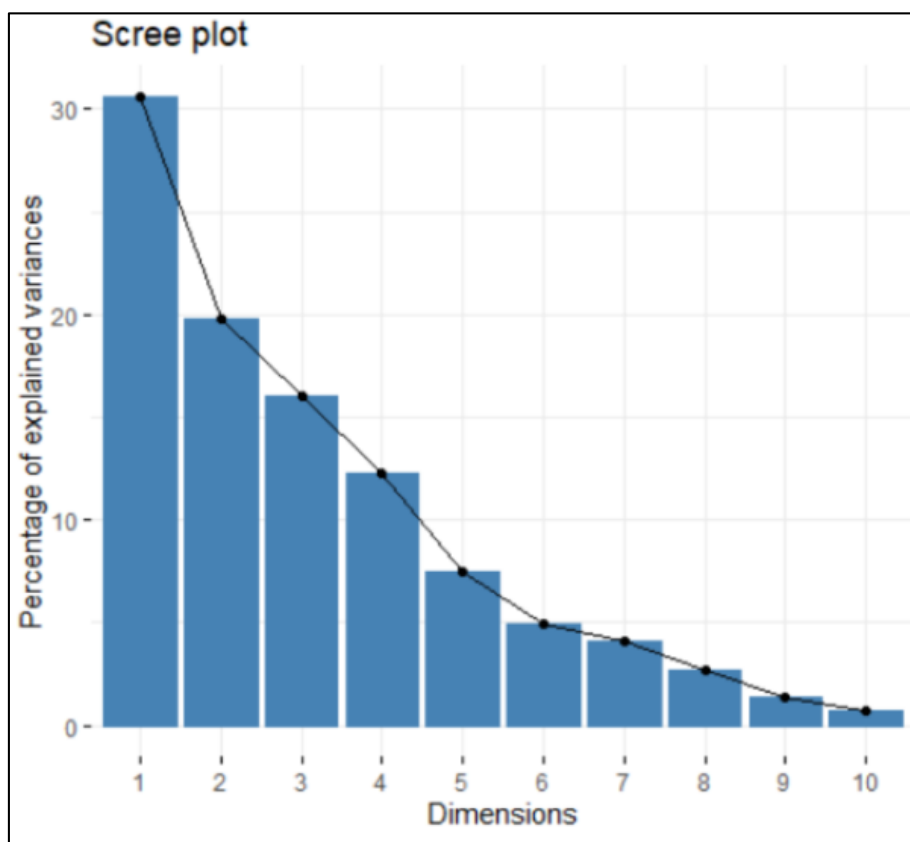
**Table 6:** PCA scores for 16 selected released varieties of sorghum.

Varieties name	PC I	PC II	PC III	PC IV
DSV-4	-1.175	0.566	2.898	0.484
SPV-86	-1.329	-3.257	1.757	-0.904
SPV-1829	-3.096	-0.946	-0.779	0.449
BJV-44	-1.098	1.020	1.338	-0.116
DSV-5	-1.956	-1.339	-2.867	1.507
A-1	-0.332	-0.094	-1.917	-3.296
Phuleyashoda	-1.966	1.346	0.330	0.143
PhuleVasudha	-0.660	3.487	-1.151	0.693
PhuleRevathi	-1.179	1.110	0.714	-0.629
M35-1 (Akola source)	-0.441	-1.361	0.064	1.038
M35-1 (Bijapur source)	1.649	0.349	0.213	-0.927
Barsi Jowar	2.812	-0.597	-0.197	1.811
Kodmurki (popular local)	2.222	-1.064	-1.302	-0.192
SVD-803	2.540	0.285	0.299	-0.349
SVD-808 lines)	2.406	-0.641	0.855	0.277
SVD-770	1.602	1.135	-0.256	0.012

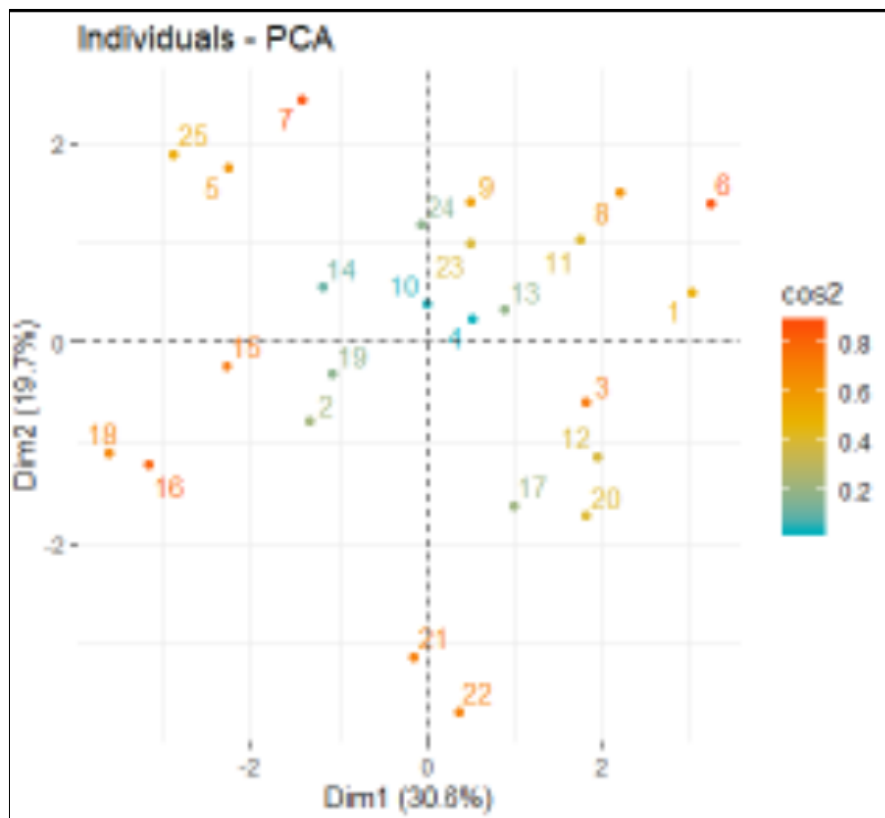
**Fig 1:** Per cent contribution of characters towards divergence of 24 selected mini core collections along with check M35-1 of *rabi* sorghum



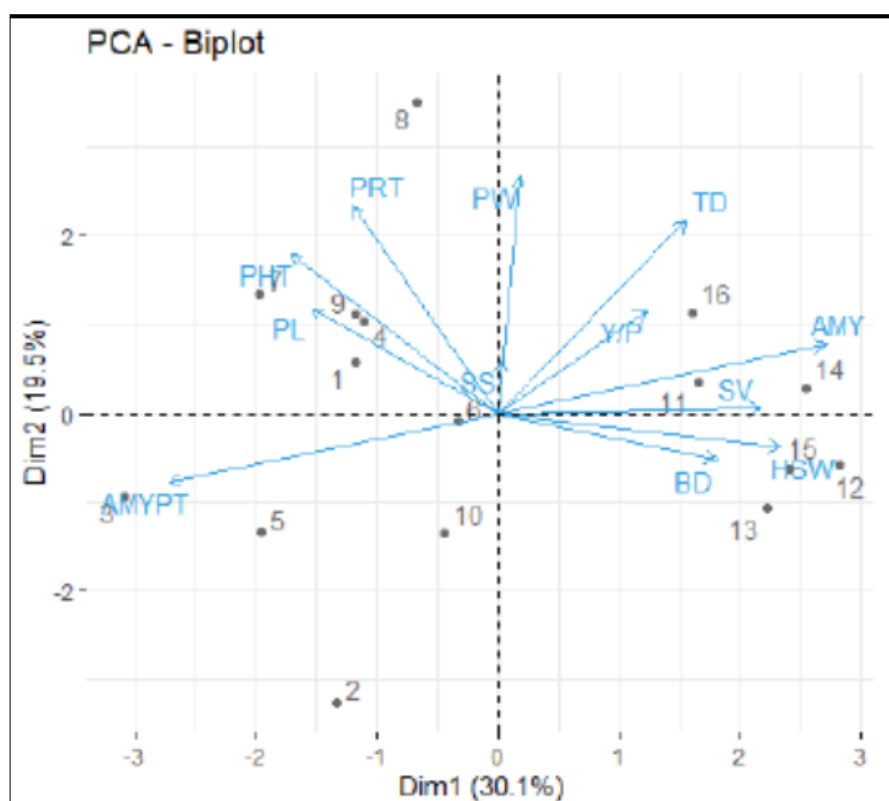
**Fig 2:** The length of arrows shows the proportion of contribution in the principal components and direction of arrows tells whether the proportion is positive or negative in 24 selected mini core collections of *rabi* sorghum.



**Fig 3:** Scree plot of data showing that maximum variability is explained by first principal component in 24 selected mini core collections of *rabi* sorghum

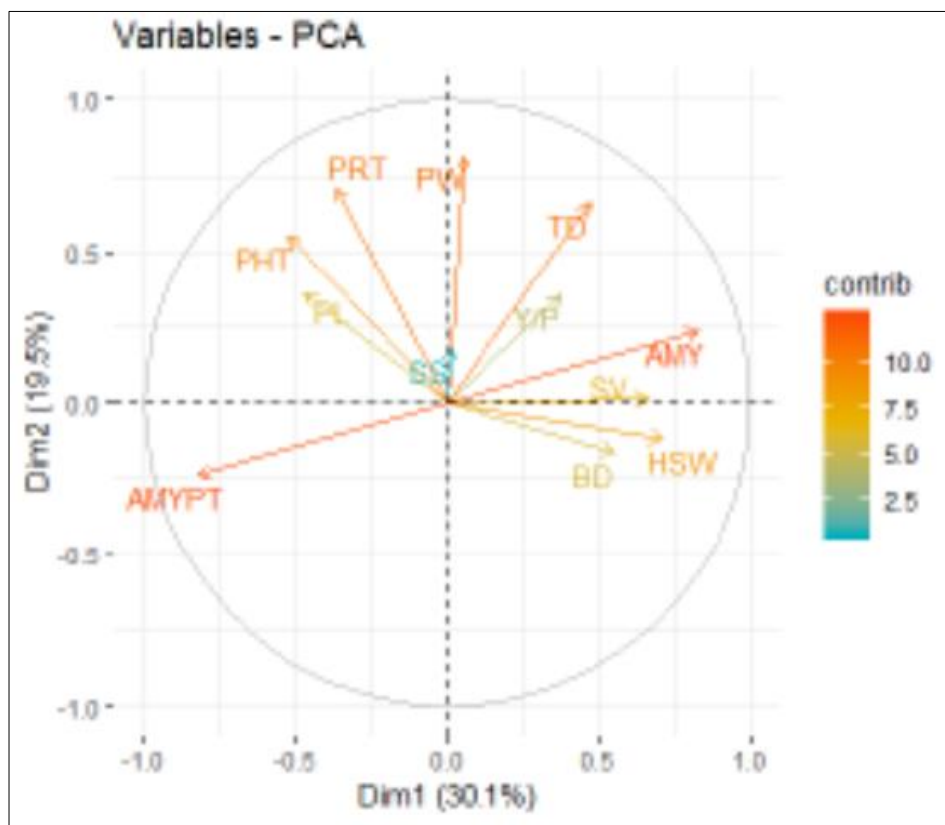


**Fig 4:** Distribution and grouping of 24 selected mini core collections of *rabi* sorghum across first two components based on PCA.

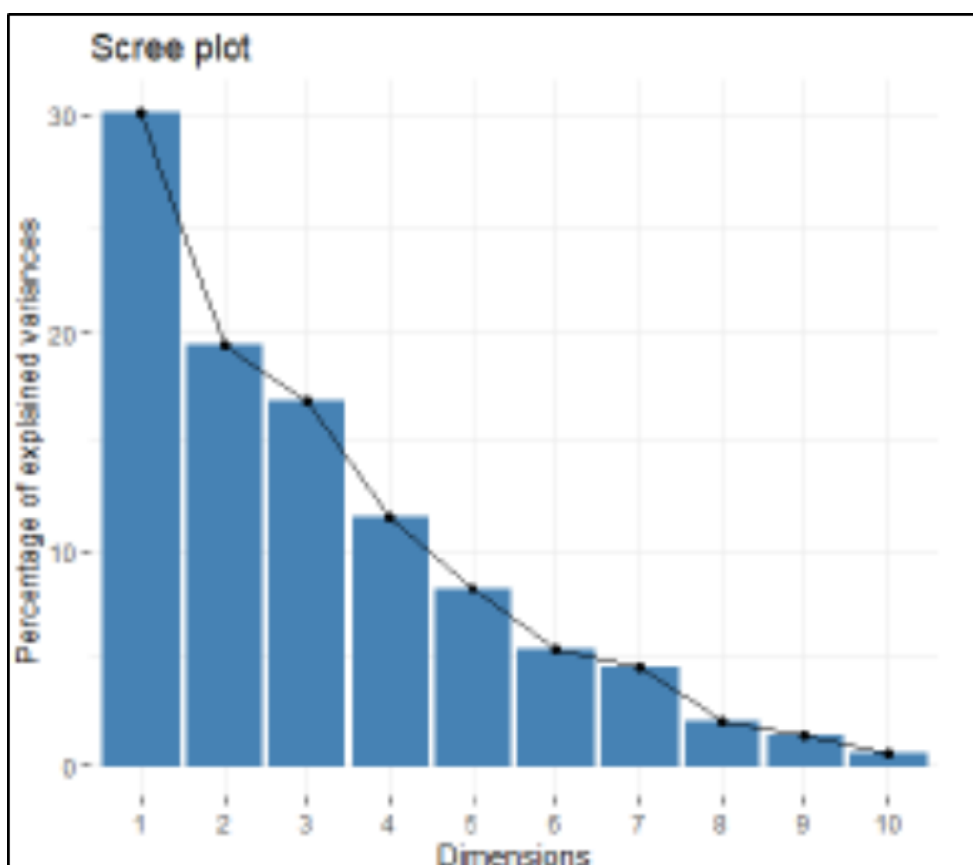


**Fig 5:** Per cent contribution of characters towards divergence 16 selected promising varieties of *rabi* sorghum.

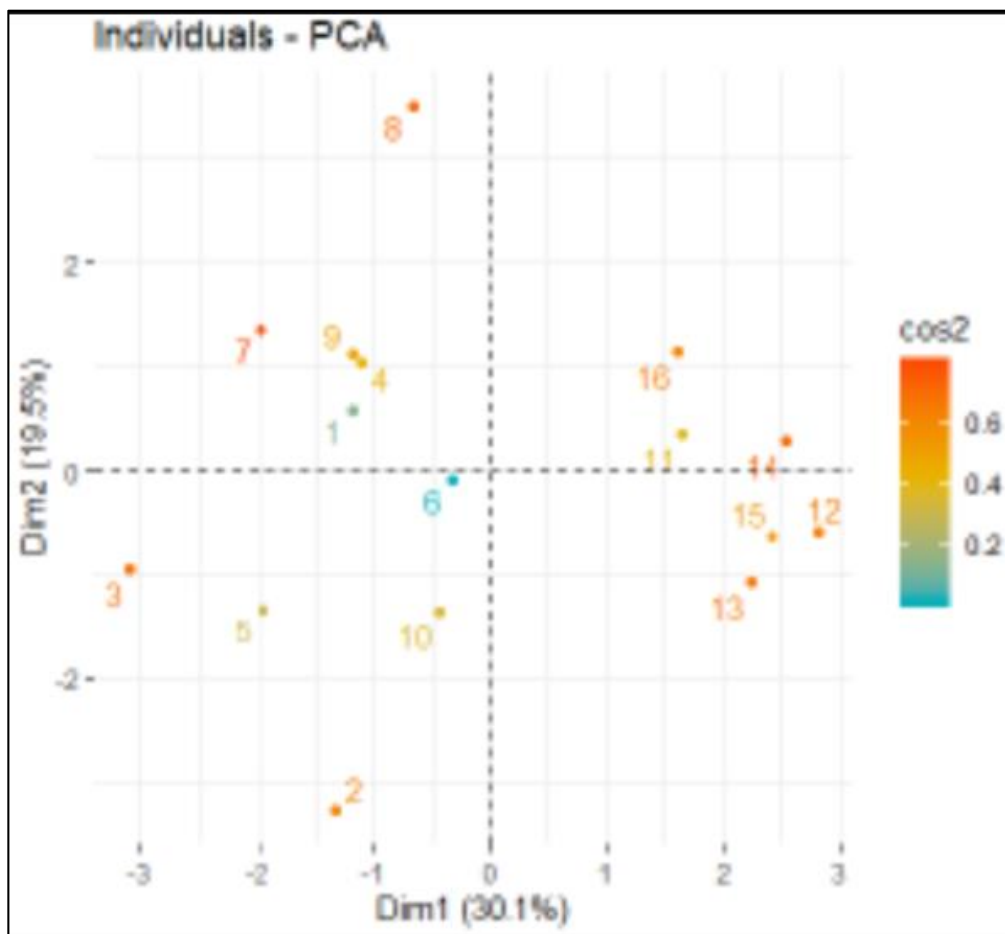




**Fig 6:** The length of arrows shows the proportion of contribution in the principal components and direction of arrows tells whether the proportion is positive or negative in selected promising varieties of *rabi* sorghum.



**Fig 7:** Scree plot of data showing that maximum variability is explained by first principal component in selected promising varieties of *rabi* sorghum



**Fig 8:** Distribution and grouping of 16 selected promising varieties of *rabi* sorghum. Across first two components based on PCA

#### Acknowledgement

The senior author is grateful to DSR Hyderabad for providing mini core collection.

#### References

1. Anonymous. World Agricultural production, 2019a, 21.
2. Anonymous. World Agricultural production, 2019b, 21.
3. Taylor JRN, Dewar J, Taylor J, Ascheraden RF. Factors affecting the porridge-making quality of South African sorghums. *Journal of the Science of Food and Agriculture*. 1997; 73:464-470.
4. Food and Agricultural Organization [FAO]. Sorghum and Millets in Human Nutrition. FAO, Rome, Italy. (FAO and Nutrition Series, 1995, 27.
5. Awika JM, Rooney LW. Sorghum phytochemicals and their potential aspects on human health. *Phytochemistry*. 2004; 65:1199-1221.
6. Dicko MH, Gruppen H, Traore AS, Alphons GJ, Voragen AGJ, Berkel WJH. Sorghum grain as human food in Africa: Relevance of content of starch and amylase activities. *African Journal of Biotechnology*. 2006; 5:384-395.
7. Mehmood S, Bashir A, Amad A, Akram Z. Molecular characterization of regional *Sorghum bicolor* varieties from Pakistan. *Pakistan Journal of Botany*. 2008; 40:2015-2021.
8. Mudjisihino R, Damardjati DS. Prospects usefulness sorghum as a source of food and feed. *Journal of Agricultural Research Development*. 1987; 1:1-5.
9. Mohammadi SA, Prasanna BM. Analysis of genetic diversity in crop plants, salient statistical tools and considerations. *Crop Sciences*. 2003; 43:235-48.
10. Johnson DE. Applied multivariate methods for data analysis. New York: Duxbury Press. 2012.
11. Das LV. Problems facing plant breeding. New Delhi: CBS, 2000.
12. Yan W, Kang MS. GGE Biplot Analysis: A graphical tool for breeders, Geneticists, and Agronomists. Boca Raton, FL: CRC Press, 2003.
13. Bello D. Genetic variability and inter-relationship of traits in local sorghum (*Sorghum bicolor* (L.) Moench) in adamawa state. Unpublished M.Sc. Thesis, Yola: Federal University of Technology, 2004.
14. Maji AT, Shaibu AA. Application of principal component analysis for ricegermplasm characterization and evaluation. *Journal of Plant Breeding and Crop Science*. 2012; 4:87-93.
15. Adams MW. An estimate of homogeneity in crop plants with special reference to genetic vulnerability in dry season. *Euphytica*. 1995; 26:665-679.
16. Sowbhagya CM, Bhattacharya KR. Simplified determination of amylose in milled rice starch. *Stärke*. 1979; 31:159-163.
17. Pearson K. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine*. 1901; 2(11):559-572. Doi: 10.1080/14786440109462720.
18. Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*. 1933; 24:417-441.
19. Chatfield C, Collins A. Introduction to Multivariate Analysis. Chapman and Hall, 1980.
20. Jolliffe IT. Principal component analysis, 2nd edn. New York, NY: Springer-Verlag, 2002.

21. Watson JW, Eyzaguirre PPB. Homegardens and in situ conservation of plant genetic resources in farming systems. Proceedings of the second international home gardens workshop, 17-19th July 2001, Witzen hausen, Federal Republic of Germany. International Plant Genetic Resources Institute, Rome, Italy, 2002.
22. Jain SK, Patel PR. Genetic Diversity and Principle Component Analyses for Fodder Yield and their Component Traits in Genotypes of Forage Sorghum (*Sorghum bicolor* L. Moench) Annals of Arid Zone. 2016; 55(1-2):17-23.
23. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A. 2016; 374:20150202.
24. Leonard K, Peter RJ. Finding Groups in Data: An Introduction to Cluster Analysis, 2009, 344.
25. Clifford HT, Stephenson W. An Introduction to Numerical Classification. Academic Press, London, 1975, 229.
26. Guei RG, Sanni KA, Fawole AFJ. Genetic diversity of rice (*O. sativa* L.). Agronomieafricaine. 2005; 5:17-28.
27. Gabriel KR. The Biplot graphic display of matrices with application to principal component analysis. Biometrika. 1971; 58(3):453-467.
28. Harris R. A multivariate analysis. Laurence Girlbaum Associates, New Jersey, USA, 2001, 487.
29. Amurrio J, De Ron A, Zeven AC. Numerical taxonomy of Iberian pea landraces based on quantitative and qualitative characters. Euphytica. 1995; 82:195-205.
30. Rabbani MA, Iwabuchi A, Murakami Y, Suzuki T, Takayanagi K. Phenotypic variation and the relationships among mustard (*Brassica juncea* L.) germplasm from Pakistan. Euphytica. 1998; 101:357-366.
31. Berdahl J, Mayland HF, Asay KH, Jeerson PG. Variation in agronomic and morphological traits among Russian wild rye accessions. Crop Science. 1999; 39:1890-1895.
32. Chandran K, Padya SM Morphological characterization of *Arachis* species of section *Arachis*. Plant Genetic Resources Newsletter. 2000; 121:38-41.
33. Cravero VP, López Anido FS, Cointy EL. Characterization selection de familles S<sub>1</sub> de alcaucil a travels de technical de analysis multivariate. Horticulture Brasilia. 2002; 20:619-625.
34. Rosso B, Pagano E. Evaluation of introduced and naturalized populations of red clover (*Trifolium pretense* L.) at Pergamino EEA- INTA, Argentina. Genetic Resources and Crop Evolution. 2005; 52:507-511.
35. Bhargava A, Shukla S, Rajan S, Ohri D. Genetic diversity for morpho-logical and quality traits in quinoa (*Chenopodium quinoa* Willd.) germplasm. Genetic Resources and Crop Evolution. 2007; 54:167-173.
36. Sharma JR. Statistical and Biometrical Techniques in Plant Breeding. New Age International (P) Limited Publishers, New Delhi, India, 1998, 432.
37. Fenty J. Analysing distances. The Strata Journal. 2004; 4:1-26.
38. GanaVarthini. Evaluation of Rice Genetic Diversity and Variability in a Population Panel by Principal Component Analysis Indian Journal of Science and Technology. 2006, 2013 and. 2014; 7(10):1555-1562.
39. Gana AS, Shaba SZ, Tsado EK. Principal component analysis of morphological traits in thirty-nine accessions of rice (*Oryza sativa*) grown in a *rainfed* lowland ecology of Nigeria. J Plant Breeding and Crop Science. 2013; 5:120-126.
40. Ayana A, Bekele E. Multivariate analysis of morphological variation in sorghum (*Sorghum bicolor* (L.) Moench) germplasm from ethiopia and eritrea. Genetic Resources and Crop Evolution. 1999; 46(3):273-84. Doi: 10.1023/A:1008657120946.
41. Ali MA, Jabran K, Awan SI, Abbas A, Zulkiffal M, Acet T. Morpho-physiological diversity and its implications for improving drought tolerance in grain sorghum at different growth stages. Australian Journal of Crop Science. 2011; 5(3):311.
42. Abraha T, Githiri SM, Kasili R, Araia W, Nyende AB. Genetic variation among sorghum (*Sorghum bicolor* (L.) Moench) landraces from Eritrea under postflowering drought stress conditions. American Journal of Plant Sciences. 2015; 6(09):1410. doi:10.4236/ajps.2015.69141.
43. Chikuta S, Odong T, Kabi F, Rubaihayo P. Phenotypic diversity of selected dual purpose forage and grain sorghum genotypes. American Journal of Experimental Agriculture. 2015; 9: 6. doi:10.9734/AJEA/2015/20577.
44. Karadi A, Kajjidoni ST. Genetic Diversity and Principal Component Analysis for Grain Quality and Productivity Traits in Sorghum [*Sorghum bicolor* (L.) Moench]. International Journal of Current Microbiology and Applied Sciences. 2019; 8(9):839-850.