ISSN 2709–3662 (Print)
ISSN 2709–3670 (Online)
https://doi.org/10.52587/JAF040205
Journal of Agriculture and Food
2023, Volume 4, No.2, pp. 50-65

Evaluation of the genetic variability between biomass and biofuel traits in *Sorghum bicolor* germplasm of Pakistan

Rehana[1], Ummara Waheed[1], Shoaib-ur-Rehman[1], Sarmad Frogh Arshad[2], Sehrish Ijaz[1*]

Abstract

Sorghum is a major fodder crop with a substantial biomass production all over the world, including Genetic divergence among 15 Pakistani sorghum genotypes was evaluated under this study. High variability was reported in fresh biomass (382.197-778.181g), dry biomass 70.717-79.288), leaf area index (51.200-69.596cm2), leaf length 24.778-34.876 cm2), plant height 188.36-290.16cm), brix value (11.703-16.034) days to 50% flowering (156.21-292.288 days), and days to maturity (109.1-120.87 days). The first five principal components (PCs) across sorghum genotypes with Eigen values >1 shared 83.25% variability. Strong positive correlation was observed among fresh and dry biomass with the number of leaves per plant, the flag leaf area index, days to maturity and 50% of the days to flowering, whereas plant height and days to maturity showed a positive correlation. The Un-weight Pair-Group Method of Analysis (UPGMA) identified 5 morphotypes. Based on homology, the germplasm was divided into five classes. The genotypes (GP-6) and (GP-7) had the highest values for the fresh biomass, stem thickness, leaf length, dry biomass, number of leaves/plants and flag leaf area index. The Pakistani sorghum germplasm's explored genetic potential may be useful for varietal development programs.

Article History: **Received:**; **Revised:**, **Accepted: 20**rh February, 2023
Keywords: Principal component analysis; UPGMA; Multivariate analysis

[1*,3,4,5] Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Shareef University of Agriculture Multan, Pakistan[2] Department of Biochemistry and Biotechnology, Faculty of Veterinary and Animal Science, Muhammad Nawaz Shareef University of Agriculture Multan, Pakistan; Corresponding author's email: sehrish.ijaz@mnsuam.edu.pk

Introduction

Sorghum is a vital crop with multiple applications, including food, feed, and biofuel (Fracasso et al., 2017). Due to natural heat and drought resistance, sorghum easily flourishes in harsh climate (Lamaoui et al., 2018). Because of its low genome size of 730 Mb and its C4 photosynthetic system, sorghum has become a model crop among tropical grasses (Pardo & VanBuren, 2021). Genetic divergence is the basis for plant improvement (Shoji et al., 2022). By using a diverse pool of genetic material, plant breeders can improve the varietal improvement program. Genetic divergence analysis is a useful tool for characterizing phenotypes and identifying the sources of reported variations and similarities in genotypes (Saba Rahim et al., 2018). It is possible to use a wide variety of molecular, morphological, and biochemical marker systems to identify and classify the many varieties of crop germplasm. The simplest and cheapest way to analyze these variations is by morphological traits (Al-Naggar et al., 2020).

The genetic variation between crop germplasms has been analyzed using a variety of various multivariate methods. Principal component analysis (PCA), principal coordinate analysis (PCoA), multidimensional scaling (MDS), and cluster analysis are some of these techniques (Dudhe et al., 2020). Principle components analysis (PCA) is a statistical method to reduce the variance of interconnected features to a small number of new variables that are statistically independent of one another (Rost & Sander, 1993). Principal component analysis (PCA) begins with the determination of the Eigenvalue, which represents the overall degree of dissimilarity along the PC axis. The first PC has the greatest amount of variety. While the first PC checks for and manages any changes that were missed, the second covers most of the changes that were related to the first (Low et al., 2003). Score plots are used to position genotypes in the coordinate system and biplot analysis is used to evaluate how well genotypes perform in various environments. Cluster analysis dendrogram displays both high similarities within clusters and large variations between them (Kasoma et al., 2021; Séralini et al., 2007). Clustering methods are divided into two categories: distance-based and model-based based (Gholamnezhad et al., 2022; Härdle & Simar, 2013). Clustering techniques that depend on distance are divided into two categories. First, there are hierarchical clustering approaches in which people who are most similar to one another are grouped together and then their relationships among themselves are used to form larger clusters (Jafarzadegan et al., 2019). Ward's minimum variance method is used and accepted after UPGMA (Videla et al., 2021; Zaida Victoria Narcisa Betancourth Aragón, 2010). Non-hierarchical approaches, such as those based on similarity or frequency of occurrence are used to classify people into distinct groups (Raychaudhuri, 1999).

The most widely grown summer feed crop in Pakistan is sorghum due to its tolerance to extreme environments. However, there is a major need for the development of sorghum genotypes for use as high-quality food and grain. To do this, it is important to keep a constant flow of different germplasm, such as landraces, introductions, wild species and relatives. Even though there are many reports about sorghum morphological diversity analysis, (Dossou-Aminon et al., 2015) this field still needs to be explored.

To date, we have screened fifteen samples of Pakistani sorghum germplasm for their high biomass potential. The length of the flag leaf was determined by measuring it from its base to its tip, while the width was determined by measuring it at three different spots along the

flag leaf, near its base, near its tip, and in its middle. The length of each leaf was calculated from its base to its tip in centimeters and its breadth was measured at its base, its mid-point, and its tip. By multiplying the length and width of the leaf, one can find the area of a leaf or the area of a flag leaf. A hand refractometer and a weighing scale were used to determine the Brix value and fresh biomass's weight in gram, respectively. The number of days to maturity (defined as the number of days from the date of sowing to the stage at which 100% of the plants had grown) and dry biomass data were obtained at the termination of the vegetative stage.

Material and Methods

Plant material and field layout: The 15 sorghum genotypes utilized in the experiments were retrieved from the Fodder Research Substation at the Ayyub Agricultural Research Institute (AARI) in Faisalabad, Pakistan. The field experiment was performed in 2021 in MNS-University of Agriculture Multan. There were three meters between each row and 75 cm distance in the plants within each row. The dibbler method was used to sow two seeds per hole for a strong plant stand. Following germination, each genotype produced fourteen plants per row, which were then trimmed such that only one plant per hole survived.

Morphological characterization: Data were collected from tagged plants per genotype and replication at 50% flowering for all parameters with the exception of days to maturity and dry biomass. Plant height was measured in centimeters from the ground to the plant's last node. The stem thickness was determined by using the Vernier Caliper. The number of days from the date of sowing to the phase when 50% of the plants flowered was recorded (Dossou-Aminon et al., 2015; Raza et al., 2020). Flag leaf width was determined at three positions, i.e., near the tip, near the base, and at the middle point of the flag leaf blade. Flag leaf length was measured from the point of origin to the tip of the flag leaf. Leaf length was measured from base to tip of each leaf in cm and leaf width was measured on three points i.e., near the base, mid, and near the tip of the leaf blade in centimeters. Leaf area and flag leaf area in dice were calculated as the product of leaf length and leaf width. The weighing balance was used to record the weight of fresh biomass and the Brix value was recorded with a hand refractometer. At the final maturation stage, the data for traits like dry biomass (measured as the number of days from the sowing date to the stage at which 100% of the plants mature) and days to maturity (recorded as the number of days from the sowing date to the stage at which 100% of the plants mature) were calculated.

Statistical analysis: The descriptive statistics (mean, SD, CV) and ANOVA in SAS 9.1 were used to evaluate the quantitative data first (L, 2013; Malghani et al., 2022; Tewodros et al., 2019). Minitab 14 was used to perform principal component analysis (PCA) on the correlation matrix and the significant loading factors (account for at least 30% of the variant) were identified (Bahrami et al., 2020; Maji A. T., 2012). Correlation coefficients between pairs of quantitative morphological characteristics were calculated using the simple Pearson method. Cluster analysis was carried out with the use of the UPGMA technique (Al-Mamun et al., 2022; Dossou-Aminon et al., 2015).

Results

Analysis of variance (ANOVA) and descriptive statistics of quantitative traits

There were significant variations across sorghum genotypes for all quantitative characteristics: Brix value, stem thickness, days to maturity, flag leaf width and area index, plant height, leaf width, fresh biomass, number of leaves/plants, and days to 50% flowering. The descriptive statistics of quantitative characters are shown in (Table 1). While plant height (188.36–290.16 cm), number of the leaf (9.16-17.02 cm), number of nodes (9.97-16.97), leaf length (24.77-34.87 cm), Leaf area index (51.2-69.59cm2), Flag leaf length (10.96-19.73cm), Flag leaf area index (15.71-33.60cm2), fresh biomass (382.19–778.18g) and dry biomass (70.71-79.8g) were more variable than the rest of the traits (Table 1). The leaf length depicted the lowest variability (1.58 % CV) among all quantitative traits. The mean values of plant height, stem girth, number of leaves per plant, number of nodes, and internodal distance. Leaf length, leaf width, leaf area index, flag leaf length, flag leaf width, flag leaf area index, fresh biomass, dry biomass, days to 50% flowering, days to maturity, and Brix value were 229.27 cm, 2.40 cm, 11.76 cm, 12.36 cm, 8.43 cm, 30.03cm, 2.79 cm, 62.75 cm2, 15.23 cm, 2.19 cm, 24.77 cm2, 530.34 g, 227.63 g, 75.21 days, 115.40 days, 13.18. Furthermore, the lowest and highest ranges for various traits are displayed in (Table 1) including fresh biomass (382.19–778.18 g) and dry biomass (70-71- 79.28 g), leaf length (24.77-34.87cm), Leaf area index (51.2-59.69 cm2).

Table 1: statistical descriptions of 15 sorghum genotypes' 16 quantitative attributes.

Traits	**Minimum**	**Maximum**	**Mean**	SD	CV (%)
PH	188.4	290.2	229.3	0.9	3.7
SG	1.9	4.2	2.4	0.5	16.4
NL	9.2	17.0	11.8	1.8	3.5
NN	9.5	16.9	12.0	1.9	16.5
INTD	6.8	11.7	8.4	1.4	4.9
LL	24.8	34.9	30.0	2.2	1.6
LW	2.5	3.1	2.8	0.2	3.6
LAI	51.2	69.6	62.7	6.3	3.8
FLL	10.9	19.7	15.2	2.5	5.7
FLW	1.9	2.5	2.2	0.2	3.9
FLAI	15.7	33.6	24.8	5.5	6.9
FB	382.2	778.2	530.3	113.7	3.2
DB	70.7	79.3	75.2	2.2	9.2
DTF	156.2	292.9	210.1	39.1	5.5
DTM	109.1	120.9	115.4	3.6	1.8
BRIX	11.7	16.0	13.2	1.1	5.3

PH - Plant height; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf width; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf, width; **FLAI** - Flag leaf area index; **FB**

-Fresh biomass; **DB** - Dry biomass; **SG** - Stem girth; **NL** - Number of leaves per plant; **NN** - Number of nodes per plants; **INTD** - Inter node distance; **DTF** - Days to 50% flowering; **DTM** - Days to maturity; **BRI** -Brix value.

Principal component analysis (PCA): The redundancy in a data collection was eliminated using PCA analysis of a number of characters (PC1, PC2, PC3, PC4, and PC5) having an Eigenvalue>117 were verified (Table 4). The Collective variability of five PCs was 83.25%. The PC1 shared 28.60% of the total variability followed by PC2 (26.15), PC3 (11.80 %), PC4 (9.37), and PC5 (7.32%). Different quantitative traits donated more than 28% to the variation factor in PC1 such as Plant height (30.3%), Stem girth (70.1%), Number of nodes per plant (47.8%), Flag leaf length (72%), Flag leaf width (37.5%), Flag leaf area index (71%), Number of leaves per plant (35.6%), Dry biomass (87.4%), Fresh biomass (85.7%), Days to 50% flowering (51.7%), Brix value (51.1%). PC1 showed a weak positive correlation with the Leaf width (0.21%). PC1 represented positive and strong factors with Fresh biomass (85.7%), Dry biomass (87.4%), Flag leaf area index (71%), Flag leaf length (72%), and Brix value (51.1%) (Table 4). PC2 contributed to 26.15% of total trait variation. While it showed a strong and positive correlation with the traits such as the number of leaves per plant (83.7%), Plant height (78.9%), Leaf area index (71.8%), Leaf length (59.3%), Number of nodes per plants (51.85%). A negative correlation was observed with the stem girth (22.2%), Flag leaf width (66.3%), Flag leaf length (43.5%), Flag leaf area index (57.5%), and Days to maturity (0.60%) in PC2 (Table 4). Brix value represented (57.5%) of the factor variation in PC2 (Table 2). The PC3 contributed to 11.80% of the total trait variation. While it showed a strong and positive correlation with the traits such as leaf width (80.8%), Leaf area index (52.6%), and Days to 50% flowering (48.4%) in PC3 (Table 2). PC3 showed a weak and negative correlation with the Number of leaves per plant (0.42%) and the number of nodes per plant (0.67%) (Table 2). The PC4 contributed to 9.37% of the total trait variation. While it exhibited a strong and positive correlation with the traits such as: Days to maturity (62.7%), Fresh biomass (41.3%), Inter node distance (38.2%), and Dry biomass (39.1%). The PC5 was associated with 7.32 % of the total trait variation. The variables Days to Maturity (56.8%) and Number of Nodes per Plant (35.8%) showed a high and positive correlation with it (Table 2). The remaining variables had weak or no discriminatory power. Thus, the most important descriptors were those associated with PC1, PC2, PC3, PC4, and PC5 (Table 4). The PC5 covered 83.25% of the total variance and traits FB, DB, NL, PH, FLL, PH and LAI shared 85.7%, 87.4%, 83.7%, 78.9%, 72%, and 71.8%, respectively of total variation (Table 2).

Table 2: Principal component analysis (PCA) of sorghum traits.

Variables	PC1	PC2	PC3	PC4	PC5
PH	0.30	0.79	-0.29	0.06	0.06
SG	0.70	-0.22	-0.34	-0.01	0.18
NL	0.35	0.84	-0.04	-0.36	0.14
NN	0.48	0.52	-0.07	-0.36	0.35

INTD	-0.54	0.39	-0.31	0.38	0.10
LL	0.28	0.59	-0.14	0.15	-0.54
LW	0.02	0.41	0.80	0.17	0.155
LAI	0.23	0.71	0.52	0.22	-0.25
FLL	0.72	-0.44	-0.23	-0.08	-0.37
FLW	0.37	-0.66	0.28	0.06	0.19
FLAI	0.71	-0.58	-0.16	-0.01	-0.16
FB	0.86	0.05	0.15	0.41	-0.02
50% DTF	0.52	-0.25	0.48	-0.45	0.22
DB	0.87	0.017	0.14	0.39	-0.02
DTM	0.15	-0.06	-0.28	0.62	0.56
BV	0.51	0.575	-0.37	-0.24	0.10
Eigenvalue	4.57	4.184	1.88	1.5	1.17
Total Variability (%)	28.6	26.15	11.8	9.37	7.32
Cumulative-variability (%)	28.6	54.75	66.55	75.93	83.25

PH - Plant height; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf width; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf width; **FLAI** - Flag leaf area index; **FB** -Fresh biomass; **DB** - Dry biomass; **SG** - Stem girth; **NL** - Number of leaves per plant; **NN** - Number of nodes per plants; **INTD** - Inter node distance; **DTF** - Days to 50% flowering; **DTM** - Days to maturity; **BRI** - Brix value.

Biplot analysis

The variables in the biplot were shown to be superimposed as vectors, with the length of each vector representing the proportion of variation in that variable. There was less resemblance between the germplasm that were plotted away from the origin and the genotypes that were plotted closer to the center. The traits like PH, NL, LAI, NN, BV, FB, LL, BV and DB were well represented and exhibited high variability in F1 and F2 axes (Figure 1). On the other hand, LW trait showed less variability especially in comparison to other traits. Quantitative traits like PH, NL, LAI, NN, BV, FB, LL, LW, NN, and DB and BV were presented at positive & positive coordinates in biplot analysis in F1 and F2 axes (Figure 1). Traits DTM, DTF, FLW, SG, FLL, and FLAI were presented at the fourth positive-negative coordinate in F1 and F2 (Figure 1). Only one trait INTD was present at the first negative-positive coordinate in biplot analysis in F1 and F2 axes (Figure 1).

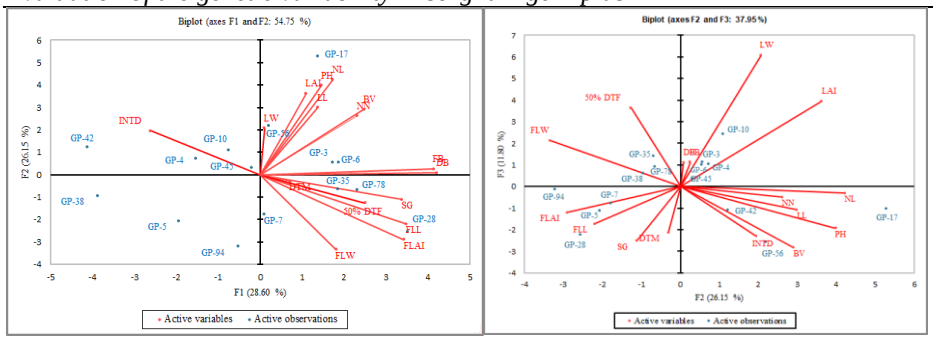

Figure (1a). Biplot analysis related to different quantitative traits in sorghum for F1 and F2 axes. Figure (1b). Biplot analysis related to different quantitative traits in sorghum in F2 and F3 axes.

The traits like LW and LAI were well represented and exhibited high variability in F2 and F3 axes (Figure 1b). The FB and DB traits showed less variability than other traits. Quantitative characters like LW, LAI, DB, and FB were allocated at a positive-positive coordinate in biplot analysis in F2 and F3 axes (Figure 1b). Traits DTF and FLW were represented at first negative-positive coordinates in F2 and F3 axes (Figure 1). Traits FLAI, FLL, SG, and DTM were allocated at a negative-negative coordinate in biplot analysis in F2 and F3 axes (Figure 1). While the traits like NL, NN, LL, PH, BV, and INTD were allocated the fourth positive-negative coordinate in biplot analysis in F2 and F3 axes (Figure 1). The characteristics such as LAI, LW, FB, and DB were all well recorded and demonstrated a significant degree of variability in both F3 and F4 axes (Figure 2). On the other hand, trait FLW presented less variability as related to other characters. Quantitative traits similar to LAI, LW, FLW, FB and DB, and BV were allocated at a positive-positive coordinate in biplot analysis in F3 and F4 axes (Figure 2). Traits DTM, INTD, LL, and PH were represented at a first negative-positive coordinate in F3 and F4 axes (Figure 2). On the other hand, Traits FLAI, FLL, SG, BV, NN, NL were allocated at a negative-negative coordinate in biplot analysis in F3 and F4 axes (Figure 2). Only DTF trait was represented at the fourth positive-negative coordinate in F3 and F4 axes (Figure 2). The trait like DTM was well represented and exhibited high variability in F4 and F5 axes (Figure 2). There was less variation in the qualities like PH, INTD, LW, and FLW, as compared to other traits. Quantitative characters like DTM, INTD, LW, FLW and PH were allocated at positive-positive coordinate in biplot analysis in F4 and F5 axes

(Figure 2). Traits like SG, BV, NL, DTF and NN were represented at first negative-positive coordinate in F4 and F5 axes (Figure 2). On the other hand, traits FLAI and FLL were allocated at negative-negative coordinate in biplot analysis in F4 and F5 axes (Figure 2). Traits like DB. FB, LL and LAI was represented at fourth positive-negative coordinate in F4 and F5 axes (Figure 2).

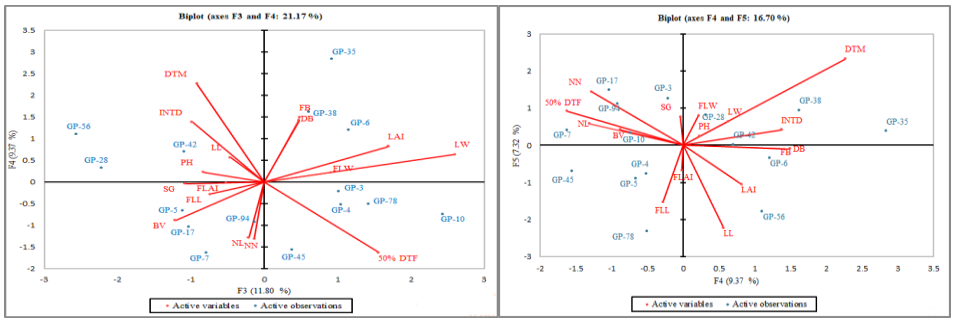

Figure 2: Biplot analysis relating to several quantitative characteristics of sorghum for F3 and F4 axes and F4 and F5 axes.

Cluster analysis

Sorghum genotypes were analyzed using UPGMA, which was generated in 5 different morphotypes for one year (Figure 3). The main cluster was divided into five subclasses, according to cluster analysis, and clusters C1 and C2 contained 11 and 4 genotypes, respectively, with 4 and 1 morphotype each (Figure 3). The centroids of the class are described and characterized in (Table 3). The genotypes (GP-7) and (GP-6) were separated as central centroids of class III and V, in that order (Table 3).

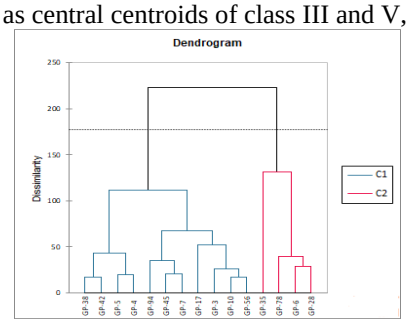

Figure 3: UPGMA cluster analysis was used to categorize 15 sorghum genotypes into 5 morphotypes.

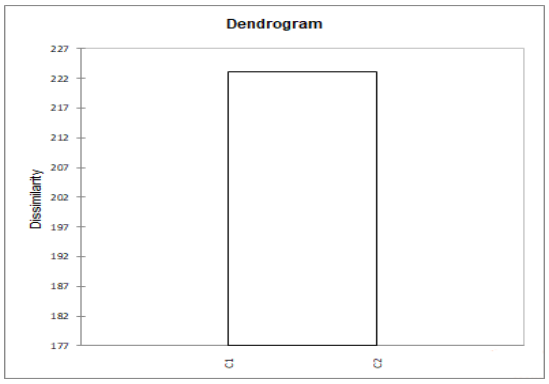

Figure 4: Classification of 15 different sorghum accessions from Pakistan based on their genetic similarity as determined by cladogenesis studies**.**

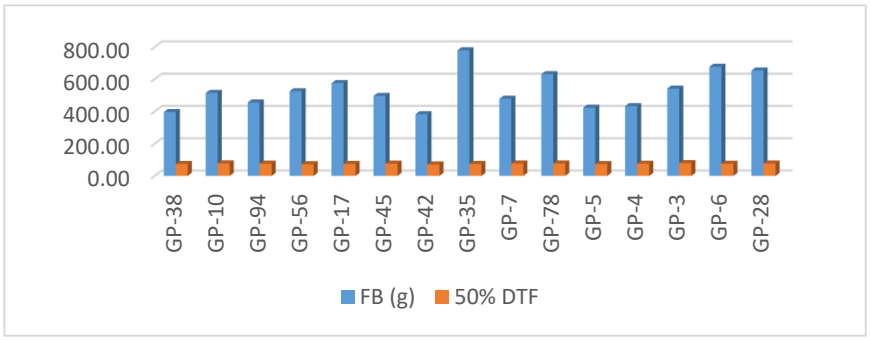

Figure 5: The 15 sorghum genotypes were compared for fresh biomass and 50% days to blooming.

Table 3. Characterization of III and V class centroids in sorghum genotypes

Cluster	PH	SG	NL	NN	INTD	LL	LW	LAI	FLL	FLW	FLAI	FB	50% DTF	DB	DTM	BV
1(GP-7)	188.36	2.479	11.624	14.613	6.793	29.640	2.468	54.852	16.209	2.176	26.457	478.162	76.520	193.713	115.992	12.499
2(GP-6)	232.56	2.520	11.619	11.625	7.310	30.148	3.078	69.596	16.818	2.120	26.733	676.177	75.162	259.357	117.038	13.798

Correlation analysis

The Pearson correlation analysis of mean value data discovered an important link between quantitative traits. PH displayed a highly significant and positive correlation with NL, NN, INTD, LL, LW, LAI, FB, DB, and BV and it shows a highly negative correlation with FLL, FLW, FLAI, and DTF (Table 4). The most of the characteristics, with the exception of FLL, FLW, FLAI, and DTF, had a lower positive phenotypic association with stem thickness (negatively correlated). NN showed a high positive phenotypic correlation with plant height (PH) and a lower positive correlation with SGNN and NL have a strong positive correlation that is extremely significant. INTD exhibits a highly significant and positive association with NL, SG, NN, and INTD, as well as a highly significant and negative correlation with PH, SG, and NL. LW showed a highly significant and positive correlation with NL, NN, INTD and a highly significant and negative correlation with SG and LL. LAI exhibited a highly significant and positive correlation with NL, LL, LW, and negatively correlated SG (Table 4). FLL exhibited a highly significant and positive correlation with SG and a high negative significance with PH, NL, NN, INTD, LW and LAI. FLW showed a highly significant and positive correlation with SG, FLL and negatively correlated with PH, NL, NN, INTD, LL and FLL. FLAI highly significant and positive correlation with SG, LW and negatively correlated FL, FLW high significant and negative correlated with PH, NL, NN, INTD, LL, and LAI (Table 4). FB showed a highly significant and positive correlation with FLAI, FLL, LAI, LW, LL, NN, NL, SG, PH and negative correlated with INTD. 50% DTF highly significant and positive correlation with FLW, SG, NL, NN, LW, FLL, FLAI, FB and negative correlated with PH, INTD and LL. DB highly significant and positive correlation with FLL, FLAI, FB, DTF, FLW, LAI, LW, LL, NN, NL, SG, PH and has negative correlation with INTD. DTM exhibited highly significant and positive correlation with SG, INTD, FB, DB and negative correlation with NL, LL, LW, LAI and FLL. BV highly significant and positive correlation with PH, NL, NN, LL, LAI, FLL, FAI, FB, DB and negative correlate with FLW in (Table 4).

4. Matrix showing correlations for 16 quantitative characteristics of sorghum.

Variables	PH	SG	NL	NN	INTD	LL	LW	LAI	FLL	FLW	FLAI	FB	50% DTF	DB
SG	0.1**													
NL	**0.8****	0.1**												
NN	0.4**	0.3**	**0.8****											
INTD	0.4**	-0.3	-0.0	-0.2										
LL	0.5**	0.0	0.5**	0.2**	0.1**									
LW	0.2**	-0.3	0.2**	0.1**	0.1**	-0.0								
LAI	0.5**	-0.1	**0.5****	0.2**	0.1**	**0.6****	**0.7****							
FLL	-0.1	**0.6****	-0.1	-0.0	-0.5	0.1**	-0.3	-0.2						
FLW	-0.2	0.3**	-0.4	-0.3	-0.4	-0.4	-0.0	-0.3	0.4**					
FLAI	-0.1	**0.5****	-0.2	-0.0	-0.5	-0.1	-0.3	-0.3	**0.9****	**0.6****				
FB	0.2**	0.5**	0.2**	0.3**	-0.4	0.3**	0.2**	0.4**	0.5**	0.3**	**0.5****			
50% DTF	-0.1	0.4**	0.1**	0.2**	**-0.6**	-0.1	0.2**	0.1**	0.3**	**0.6****	0.4**	0.2**		
DB	0.2**	0.5**	0.1*	0.3**	-0.4	0.3**	0.2**	0.3**	**0.5****	0.3**	**0.6****	**0.9****	0.2**	
DTM	0.1*	0.3**	-0.1	0.1*	0.1**	-0.0	-0.1	-0.1	-0.0	0.1	0.0	0.3**	-0.1	0.3**
BV	**0.8****	0.3**	**0.8****	**0.5****	0.0	0.3**	0.0	0.2**	0.3**	-0.2	0.1**	0.3**	0.0	0.3**

Ns = non-significant, *****=Significant <0.05, ******=High Significant <0.01, probability level, *******=High Significant at >0.0001 probability level.

PH - Plant height; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf width; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf width; **FLAI** - Flag leaf area index; **FB** - Fresh biomass; **DB** - Dry biomass; **SG** - Stem girth; **NL** - Number of leaves per plant; **NN** - Number of nodes per plants; **INTD** - Inter node distance; **DTF** - Days to 50% flowering; **DTM** - Days to maturity; **BRI** - Brix value.

Score plot analysis:

The scatter plot was used to analyze 16 traits in sorghum genotypes. The genotypes were classified into three categories (G1, G2, and G3) using principal component analysis (PCA) (Figure 6). Only four GP-10, GP-4, GP42 and GP-45 genotypes were grouped in G1. Whereas, G2 gathered four GP-17, GP-56, GP-3, and GP-6 genotypes based on DTM, BV, FLL, FB, DB, PH, DTF, and LL. Group G3 assembled four GP-7, GP-35, genotypes based on, ST, LAI, NL, FLW, LW and FLAI (Figure 6). The three genotypes GP-38, GP-5, and GP-94 were grouped in the G0 group (negative coordinate of the coordinate plot). In the coordinate system, the genotypes that were plotted further away from the central point were regarded to be more variable than those plotted closer to the central point. The GP-17, GP-56, GP-3, and GP-6 genotypes were the most promising in G2 (Figure 6). These genotypes may have a higher level of resistance to both biotic and abiotic stresses. Three of the most diverse genotypes were found in G0: GP-38, GP-5, and GP-94. These genotypes can be regarded as also being resistant to biotic and abiotic stressors. The GP-42, GP-4, and GP-10 were among the future genotypes with higher biomass characteristics found in the G1 group, whereas the GP-17, GP-56, GP-28, GP-7, and GP-78 were genotypes from the G2 and G3 groups (Figure 6).

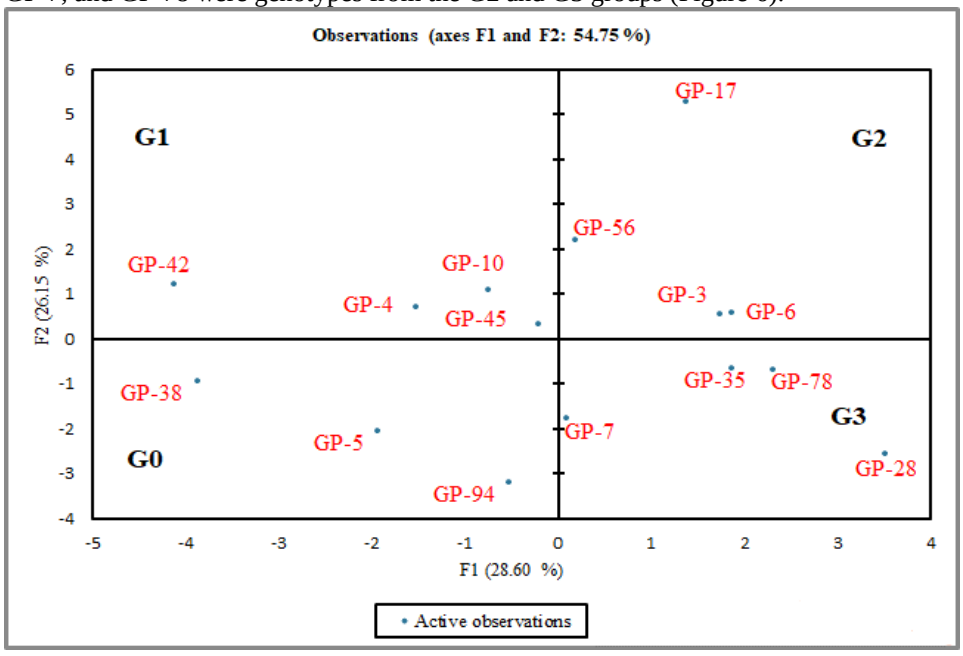

Figure 6: CA grouping of 15 varieties of sorghum from Pakistan using discriminatory quantitative characteristics

Identification of desirable genotypes for the sorghum improvement program: Two genotypes were identified for the sorghum improvement program based on morphological characterization, evaluating the financial advantage characteristics (days until blooming and fresh biomass) of Pakistani farmers. In principle component analysis, the selected genotypes were identified in positive coordinates. The two genotypes (GP-6, GP-7)

screened out as a class centroid from the UPGMA analysis. Their fresh biomass ranged from 478.162 to 676.177g; the top value 6.184 to 7.637cm was recorded in (GP-7) and (GP-6) respectively. These genotypes are known as best performers for the production of biomass, and biofuels and can be further to a national breeding program.

Discussion

Morphological classification is a very important first step in analyzing and classifying a wide range of genetic diversity present in the existing genotypes (Al-Naggar et al., 2020; Murray et al., 2009; Using, 2023). The multivariate analysis tools were used to evaluate the Pakistani sorghum germplasm in the current study. As previously reported by (Raza et al., 2020), our current results showed that each of the sixteen quantitative traits was highly significant. The sorghum germplasm's range figures for days to plant height, flowering, leaf area index, number of fresh leaves, biomass, and dry biomass are similar to previous research (Emendack et al., 2021). The majority of Pakistan sorghum genotypes with intermediate height, leaf number, and leaf area index had higher biomass as reported in Ethiopian sorghum landraces (Firezer et al., 2020). We believe that Pakistan's various sorghum germplasm is different from the Ethiopian sorghum germplasm. So these different genotypes may be beneficial for variety development (Bhusal et al., 2013).

It is easier to discover morphological differences between and across germplasms that have several variates statistical tools (Islam et al., 2018; Sümbül et al., 2022). The five PCs contributed 83.25% variation of overall diversity among the genotypes in the current study. About 28.60% variation of the total variance was found by PC1. Among 15 sorghum genotypes, PC2, PC3, PC4, and PC5 are the highest contributors having a share of 26.15%, 11.80%, 9.37% and 7.32% variation in the total variability respectively. Genotypes with high phenotypic variation on the first axis produce more biomass. The third axis can contribute into the biofuel breeding strategy (Maji A. T., 2012; Muhammad et al., 2020). Due to their effective selective strength about the analyzed characteristics and maximal diversity among them, the genotypes far away from the origin (GP-17, GP-56, GP-3, and GP-6) in G2 can be used in heterosis breeding programs. Generally, genotypes are screened from the five PCs. Mini core collection for sorghum in Pakistan can be determined by grouping the genotypes (Azameti et al., 2020). We discover a positive correlation between a number of variables such as plant height, days until 50% of the plant has flowered, and Brix value. These related traits are important for the selection of genotypes with high biofuel content. In earlier studies, a similar correlation pattern was reported (Abubakar & Bubuche, 2013; Aftab et al., 2020).

In the current study, 5 morphotypes of sorghum genotypes were yielded with help of UPGMA analysis. To increase the diversity of a breeding program, two elite genotypes (GP-7 and GP-6) that belonged to separate clusters or morphotypes can be crossed (Arshad et al., 2017). These genotypes can undergo mass selection for three generations to boost homozygosity of the desired quantitative characteristics (Alahmad et al., 2018; Qu, 2007). Because of their improved climatic stress tolerance, fresh biomass, and earliness, these genotypes can be directly produced to enhance sorghum yield. Additionally, as previously suggested by Dissou-Aminon et al., (2015), the two genotypes with high fresh biomass (GP-7 and GP-6) and early maturing genotypes (GP-10, GP-94, GP-17, GP-56, GP-45, GP-78, and GP-5) can be crossed to produce massive biomass and early maturing varieties.

Conclusion

The current study estimates the largest genotype collection of Pakistani sorghum to date for the evaluation of phenotypic variation. The best performing genotypes related to biomass and biofuels are potentially identified in this study which will improve the present sorghum breeding strategy. For association mapping, principal component analysis can be useful for a variety of genotype sets. The five classes were divided into class nominators by the cluster analysis, which can give the novel material with built-in tolerance to biotic and abiotic stress. In multi-location trials, the two genotypes GP-6 and GP-7 can be grown directly.

Reference

Abubakar, L. & Bubuche, T.S. (2013). *Correlation analysis of some agronomic traits for biomass improvement in sorghum (Sorghum Bicolor L . Moench) genotypes in North-Western Nigeria*. *8*(28), 3750–3756.

Aftab, M., Fatima, S., Qamar, H., Hassan, M., Zubair, M., Arshad, A., Mahmood, T., Ali, S., Shahazad, S. & Khurshid, M.R. (2020). Study of Morphological Characters Give an Insight into the Genetic Variation Present in Brassica napus L. Germplasm. *Life Science Journal*, 17(3), 56–61.

Al-Mamun, M., Rafii, M., Oladosu, Y., Misran, A.B., Berahim, Z., Ahmad, Z., Arolu, F. & Khan, M.H. (2022). Genetic Diversity among Kenaf Mutants as Revealed by Qualitative and Quantitative Traits. *Journal of Natural Fibers*, 19(11), 4170–4187.

Al-Naggar, A. M.M., Shafik, M.M. & Musa,Y.M. (2020). Genetic Diversity Based on Morphological Traits of 19 Maize Genotypes Using Principal Component Analysis and GT Biplot. *Annual Research & Review in Biology*, 35(2), 68–85.

Alahmad, S., Dinglasan, E., Leung, K.M., Riaz, A., Derbal, N., Fels, K. P.V., Able, J.A., Bassi, F.M., Christopher, J. & Hickey, L.T. (2018). Speed breeding for multiple quantitative traits in durum wheat. *Plant Methods*, 1–15.

Arshad, S.F., Sadia, B., Awan, F.S. & Jaskani, M.J. (2017). Estimation of genetic divergence among sorghum germplasm of Pakistan through multivariate tools. *International Journal of Agriculture and Biology*, 19(5), 1099–1106.

Azameti, M.K., Vishalakshi, B., Umakanth, B., Balram, M., Prasad, M.S. & Madhav, M.S. (2020). Molecular characterization of popular rice (Oryza sativa L .) varieties of India and association analysis for blast resistance. *Genetic Resources and Crop Evolution*, 67(8), 2225–2236.

Bahrami, M., Khaksar, E. & Khaksar, E. (2020). Spatial variation assessment of groundwater quality using multivariate statistical analysis (case study: Fasa plain, Iran). *Journal of Groundwater Science and Engineering*, 8(3), 230–243.

Bhusal, N., Pahuja, S.K. & Dharmveer. (2013). Correlation and path coefficient analysis for some DUS traits in forage sorghum [Sorghum bicolor (L.) Moench] genotypes. *Annals of Biology*, 29(2), 127–131.

Dossou-Aminon, I., Loko, L.Y., Adjatin, A., Ewédjè, E.E.B.K., Dansi, A., Rakshit, S., Cissé, N., Patil, J.V., Agbangla, C., Sanni, A., Akoègninou, A. & Akpagana, K. (2015). Genetic divergence in northern benin sorghum (Sorghum bicolor L. Moench) landraces as revealed by agromorphological traits and selection of candidate genotypes.

Scientific World Journal, 1-5.

Dudhe, M.Y., Mulpuri, S., Meena, H.P., Ajjanavara, R.R.G., Kodeboyina, V.S. & Adala, V.R. (2020). Genetic Variability, Diversity and Identification of Trait-Specific Accessions from the Conserved Sunflower Germplasm for Exploitation in the Breeding Programme. *Agricultural Research*, 9(1), 9–22.

Emendack, Y., Sanchez, J., Hayes, C., Nesbitt, M., Laza, H. & Burke, J. (2021). Seed-to-seed early-season cold resiliency in sorghum. *Scientific Reports*, 11(1), 1–12.

Firezer, G., Firew, M., Taye, T., Temesgen, M. & Kassahun, B. (2020). Phenotyping sorghum [Sorghum bicolor (L.) Moench] for drought tolerance with special emphasis to root angle. *African Journal of Agricultural Research*, 16(8), 1213–1222.

Fracasso, A., Perego, A. & Amaducci, S. (2017). Characterisation of ten commercial sorghum genotypes grown under water-limited conditions for bioenergy production in mediterranean environment. *Italian Journal of Agronomy*, 12(4), 302–314.

Gholamnezhad, P., Broumandnia, A. & Seydi, V. (2022). A model-based many-objective evolutionary algorithm with multiple reference vectors. *Progress in Artificial Intelligence*, 11(3), 251–268.

Härdle, W. K. & Simar, L. (2013). Applied multivariate statistical analysis. *Applied Multivariate Statistical Analysis*, 1, 1–516.

Islam, M.Z., Khalequzzaman, M., Bashar, M.K., Ivy, N.A., Mian, M.A. K., Pittendrigh, B.R., Haque, M.M. & Ali, M.P. (2018). Variability Assessment of Aromatic Rice Germplasm by Pheno-Genomic traits and Population Structure Analysis. *Scientific Reports*, 8(1), 1–14.

Jafarzadegan, M., Safi-Esfahani, F. & Beheshti, Z. (2019). Combining hierarchical clustering approaches using the PCA method. *Expert Systems with Applications*, 137, 1–10.

Kasoma, C., Shimelis, H., Laing, M.D., Shayanowako, A. I.T. & Mathew, I. (2021). Revealing the genetic diversity of maize (Zea mays L.) populations by phenotypic traits and DArTseq markers for variable resistance to fall armyworm. *Genetic Resources and Crop Evolution*, 68(1), 243–259.

L, O.F.G.H. (2013). Estimation of genetic diversity for CLCuV, earliness and fiber quality traits using various statistical procedures in different crosses of Gossypium hirsutum L. *Вестник Орловского Государственного Аграрного Университета*, *43*(4), 2–9.

Lamaoui, M., Jemo, M., Datla, R. & Bekkaoui, F. (2018). Heat and drought stresses in crops and approaches for their mitigation. *Frontiers in Chemistry*, 6(6), 1–14.

Low, C., Hsu, C.M. & Yu, F.J. (2003). Analysis of variations in a multi-variate process using neural networks. *International Journal of Advanced Manufacturing Technology*, 22(11–12), 911–921.

Maji A.T. (2012). Application of principal component analysis for rice germplasm characterization and evaluation. *Journal of Plant Breeding and Crop Science*, 4(6), 87–93.

Malghani, A.L., Amin, M. & Aqeel, M. (2022). Assesment of genetic diversity on the base of morphological traits among elite genotypes of watermelon*(Citrullus SP .)*. Pakistan Journal of Phytopathology, 23(2), 159-165.

Muhammad, A., Hao, H., Xue, Y., Alam, A., Bai, S., Hu, W., Sajid, M., Hu, Z., Abdul, R., Li, Z., Liu, P., Gong, Z., Wang, L., Liu, Z. L.Á. P. & Wang, G.Á.L. (2020). Survey

of wheat straw stem characteristics for enhanced resistance to lodging. *Cellulose*, 27(5), 2469–2484.

Murray, S.C., Rooney, W.L., Hamblin, M.T., Mitchell, S.E. & Kresovich, S. (2009). Sweet Sorghum Genetic Diversity and Association Mapping for Brix and Height. *The Plant Genome*, 2(1).

Pardo, J., & VanBuren, R. (2021). Evolutionary innovations driving abiotic stress tolerance in C4 grasses and cereals. *Plant Cell*, 33(11), 3391–3401.

Qu, X. (2007). Multivariate Data Analysis. *Technometrics*, 49(1), 103–104.

Raychaudhuri, S. (1999). DNA microarrays measure the expression of thousands of genes simultaneously , and have been described elsewhere. Arab Journal of Plant Protection, *Lander*, 455–466.

Raza, M.K., Jaskani, M.J., Naqvi, S.A. & Awan, F.S. (2020). Exploitation of Phenotypic Diversity in Male Accessions of Date Palm (Phoenix dactylifera) and its Use in Germplasm Conservation. *International Journal of Agriculture and Biology*, 24(1), 133–144.

Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. In *Journal of Molecular Biology* (Vol. 232, Issue 2, pp. 584–599).

Saba Rahim, M., Sharma, H., Parveen, A. & Roy, J.K. (2018). Trait mapping approaches through association analysis in plants. *Advances in Biochemical Engineering/Biotechnology*, 164, 83–108.

Séralini, G.E., Cellier, D. & De Vendomois, J.S. (2007). New analysis of a rat feeding study with a genetically modified maize reveals signs of hepatorenal toxicity. *Archives of Environmental Contamination and Toxicology*, 52(4), 596–602.

Shoji, T., Umemoto, N. & Saito, K. (2022). Genetic divergence in transcriptional regulators of defense metabolism: insight into plant domestication and improvement. *Plant Molecular Biology*, 109(4–5), 401–411.

Sümbül, A., Yildiz, E. & Nadeem, M.A. (2022). Elucidating the genetic variations among Turkish grape varieties using morphological and molecular markers. *Genetic Resources and Crop Evolution*, *1*(23)*45-67*.

Tewodros, B.N., Fredah, K.R.W., Wassu, M.A., Githiri, S.M. & Willis, O.O. (2019). Phenotype characterization and diversity assessment of mango (Mangifera indica L.) cultivars in Ethiopia. *Journal of Plant Breeding and Crop Science*, 11(2), 55–67.

Videla, M.E., Iglesias, J. & Bruno, C. (2021). Relative performance of cluster algorithms and validation indices in maize genome-wide structure patterns. *Euphytica*, 217(10).

Zaida V.N.B A. (2010). Isolation of proteins from Datura stramonium has ability to inhibition the multiplication of potato virus Y (PVYn). Jerash Journal of Research Studies, 58(301), 236–244.

Citataion of article

Rehana, Waheed, U., Rehman, S.U., Arshad, S.F., and Ijaz, S. (2023). Evaluation of the genetic variability between biomass and biofuel traits in *Sorghum bicolor* germplasm of Pakistan. *Journal of Agriculture and Food.* 3(2), 50-65