# Evaluation of the Genetic Variability Between Biomass and Biofuel Traits through Multivariate tools in *Sorghum bicolor* Germplasm of Pakistan

# Rehana1, Sehrish Ijaz1,\*, Sarmad Frogh Arshad, Shoaib-ur-Rehman1, Ummara Waheed1

# 1, Institute of Plant Breeding and Biotechnology, Muhammad Nawaz Shareef University of Agriculture Multan, Pakistan

# Corresponding author’s email: sehrish.ijaz@mnsuam.edu.pk

# Abstract

# Sorghum is a significant fodder crop with a high production for biomass production all around the world, including Pakistan. By analyzing sixteen different quantitative parameters over the duration of one-year, genetic divergence among 15 Pakistani sorghum genotypes was evaluated. 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). Among sorghum genotypes, first five principal components (PCs) with Eigen value>1 shared 83.25% variability. Positive correlation was observed between plant height and days to maturity, whereas fresh and dry biomass had significant positive correlation with leaf area index, number of leaves per plant, flag leaf area index, days to maturity and 50% days to flowering. 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 stem thickness, leaf length, dry biomass, fresh biomass, flag leaf area index and number of leaves/plants. The Pakistani sorghum germplasm's explored genetic potential may be useful for varietal development programs. Principle component analysis can also be used to screen the diverse set genotypes.

# Keyword: Principal component analysis; UPGMA; Multivariate analysis; Sorghum

## 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 (Kausar *et al.* 2014). Because of its low genome size of 730 Mb and its C4 photosynthetic system, sorghum has become a model crop among tropical grasses (Pardo and Van Buren 2021). Genetic divergence is the basis for plant improvement (Mohammad *et al.* 2014). 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 (Nawaz *et al.* 2004; 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 (Rakshit *et al.* 2012).

Several different multivariate approaches are being used to study the genetic deviation between crop germplasms. These methods include multidimensional scaling (MDS), Principal Component Analysis (PCA), principal coordinate analysis (PCoA), and cluster analysis (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 (Wiley 1981). 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 (Meulman 1992). 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 (Hair Jr *et al.* 1995; Kasoma *et al.* 2021). Clustering methods are divided into two categories: distance-based and model-based (Johnson and Wichern 1992). 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 (Ward Jr 1963; Videla *et al.* 2021). Non-hierarchical approaches, such as those based on similarity or frequency of occurrence are used to classify people into distinct groups (Everitt and Dunn 1992).

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. The area of a leaf or the area of a flag leaf was determined by multiplying the length and width of the leaf. The fresh biomass was measured in grams using a weighing scale and the Brix value was recorded using a hand refractometer. Days to maturity data (recorded as the number of days from sowing date to stage when 100% of plants get matured) and dry biomass data (sun-dried single plant samples were weighed in grams with weighing balance) were computed at the end of the growing cycle.

## Material and Methods

## Plant material and field layout

The 15 sorghum genotypes used in the experiments were collected from the Fodder Research Substation at the Ayyub Agricultural Research Institute (AARI) in Faisalabad, Pakistan (Table 1.1). 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 two seeds per hole for a good plant stand were sown by the dibbler method. After germination, there were fourteen plants per row of each genotype and they were thinned such that one plant was kept per hole.

|  |  |  |
| --- | --- | --- |
| **Sr. No** | **Genotypes** | **Source** |
| 1 | GP-38 | AARI |
| 2 | GP-10 | AARI |
| 3 | GP-94 | AARI |
| 4 | GP-56 | AARI |
| 5 | GP-17 | AARI |
| 6 | GP-45 | AARI |
| 7 | GP-42 | AARI |
| 8 | GP-35 | AARI |
| 9 | GP-07 | AARI |
| 10 | GP-78 | AARI |
| 11 | GP-05 | AARI |
| 12 | GP-o4 | AARI |
| 13 | GP-03 | AARI |
| 14 | GP-06 | AARI |
| 15 | GP-28 | AARI |

**Table 1.1: List of sorghum genotypes used in multivariate analysis.**

## Morphological characterization

Data were collected from tagged plants per genotype and replication at 50% flowering for all parameters except 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.* 2019. 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. The data for traits such as dry biomass (sundried single plant samples were weighed in grams with weighing balance) and days to maturity (recorded as the number of days from the sowing date to the stage when 100% of plants get matured) were calculated at the final maturation stage.

## Statistical analysis

The descriptive statistics (mean, SD, CV) and ANOVA in SAS 9.1 were used to evaluate the quantitative data first (Arshad *et al.* 2017). 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 (Maji and Shaibu 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 (Swofford and Olson 1990: Dossou-Aminon *et al.* 2015).

## Results

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

There are significant variations across sorghum genotypes for all quantitative characteristics: stem thickness, Brix value, days to maturity, leaf length, flag leaf length, plant height, leaf width, leaf area index, fresh biomass, number of leaves/plants, days to 50% flowering, flag leaf area index and flag leaf width (Table 1.2). The descriptive statistics of quantitative traits are shown in (Table 1.3). 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 1.3). 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. The lowest and highest ranges for different traits are also shown in (Table 2.2) 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 1.3).

**Table 1.3: Descriptive statistics for 16 quantitative traits of 15 sorghum genotypes.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Traits** | **Minimum** | **Maximum** | **Mean** | **SD** | **CV (%)** |
| **PH** | 188.36 | 290.16 | 229.27 | 0.888 | 3.69 |
| **SG** | 1.872 | 4.231 | 2.398 | 0.552 | 16.37 |
| **NL** | 9.168 | 17.022 | 11.85 | 1.811 | 3.54 |
| **NN** | 9.472 | 16.97 | 12.05 | 1.929 | 16.48 |
| **INTD** | 6.793 | 11.735 | 8.426 | 1.446 | 4.97 |
| **LL** | 24.778 | 34.876 | 30.03 | 2.167 | 1.58 |
| **LW** | 2.468 | 3.078 | 2.785 | 0.212 | 3.65 |
| **LAI** | 51.2 | 69.596 | 62.75 | 6.326 | 3.80 |
| **FLL** | 10.961 | 19.738 | 15.23 | 2.544 | 5.73 |
| **FLW** | 1.895 | 2.513 | 2.189 | 0.162 | 3.90 |
| **FLAI** | 15.717 | 33.603 | 24.77 | 5.477 | 6.99 |
| **FB** | 382.197 | 778.181 | 530.3 | 113.7 | 3.23 |
| **DB** | 70.717 | 79.288 | 75.21 | 2.227 | 9.19 |
| **DTF** | 156.211 | 292.866 | 210.1 | 39.1 | 5.54 |
| **DTM** | 109.104 | 120.872 | 115.4 | 3.626 | 1.86 |
| **BRIX** | 11.703 | 16.034 | 13.18 | 1.077 | 5.26 |

**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.

**Table 1.2: Analysis of variance (ANOVA) of sorghum accessions for morphological traits.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Traits** | **PH** | **SG** | **NL** | **NN** | **INTD** | **LL** | **LW** | **LAI** | **FLL** | **FLW** | **FLAI** | **FB** | **DB** | **DTF** | **DTM** | **BRIX** |
| **Rep** | 2 | 0.49 | 1.27 | 8.248 | 2.62 | 3.51 | 0.51 | 391.97 | 2.09 | 0.18 | 38.6 | 85 | 400 | 63.44 | 8.01 | 0.18 |
| **Gen** | 236.1\*\*\* | 0.91\*\* | 10.8\*\*\* | 11.16\*\*\* | 6.27\*\* | 14.1\*\* | 0.14\*\* | 120.04\*\*\* | 19.41\*\*\* | 0.08\* | 89.97\*\*\* | 38784\*\*\* | 4000\*\*\* | 14.88\*\*\* | 39.46\*\*\* | 3.48\*\*\* |
| **Error** | 0.08 | 0.15 | 0.16 | 3.93 | 0.18 | 0.22 | 0.01 | 5.69 | 0.76 | 0.01 | 3 | 292.8 | 437.4 | 17.35 | 4.59 | 0.48 |
| **Total** | 15 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

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

**Rep** - replication; **Gen** - genotype; **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)

PCA analysis of several characters was used to eradicate the redundancy in a data set. Five principal components (PC1, PC2, PC3, PC4, and PC5) having an Eigenvalue>117 were verified (Table 1. 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 2.3). 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 1.4). Brix value represented (57.5%) of the factor variation in PC2 (Table 1.4). 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 1.4). 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 1.4). 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 contributed to 7.32% of the total trait variation. While it exhibited a strong and positive correlation with the traits such as: Days to maturity (56.8%) and Number of nodes per plant (35.8%) in (Table 1.4). 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 2.3). 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 1.4).

### Table 1.4: Principal component analysis (PCA) of sorghum traits.

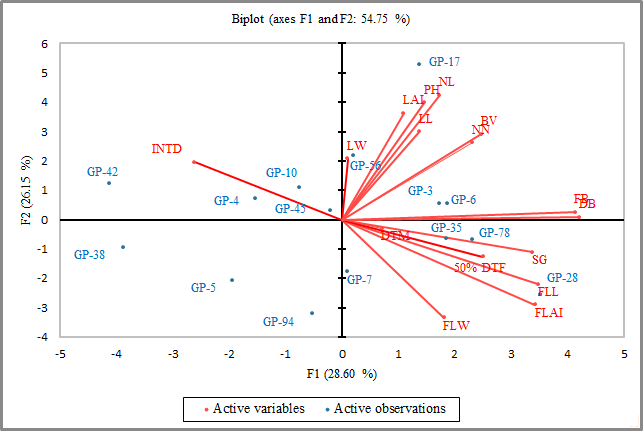
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variables** | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** |
| PH | 0.303 | 0.789 | -0.258 | 0.063 | 0.061 |
| SG | 0.701 | -0.22 | -0.336 | -0.01 | 0.185 |
| NL | 0.356 | 0.837 | -0.042 | -0.36 | 0.142 |
| NN | 0.478 | 0.518 | -0.067 | -0.36 | 0.353 |
| INTD | -0.54 | 0.386 | -0.306 | 0.382 | 0.102 |
| LL | 0.283 | 0.593 | -0.145 | 0.157 | -0.544 |
| LW | 0.021 | 0.409 | 0.808 | 0.176 | 0.155 |
| LAI | 0.227 | 0.71 8 | 0.526 | 0.226 | -0.257 |
| FLL | 0.72 | -0.44 | -0.231 | -0.08 | -0.376 |
| FLW | 0.375 | -0.66 | 0.283 | 0.06 | 0.196 |
| FLAI | 0.71 | -0.58 | -0.16 | -0.01 | -0.166 |
| FB | 0.857 | 0.05 | 0.15 | 0.413 | -0.023 |
| 50% DTF | 0.518 | -0.25 | 0.484 | -0.45 | 0.225 |
| DB | 0.874 | 0.017 | 0.147 | 0.391 | -0.026 |
| DTM | 0.149 | -0.06 | -0.285 | 0.627 | 0.568 |
| BV | 0.511 | 0.575 | -0.377 | -0.24 | 0.103 |
| Eigenvalue | 4.577 | 4.184 | 1.888 | 1.5 | 1.172 |
| 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

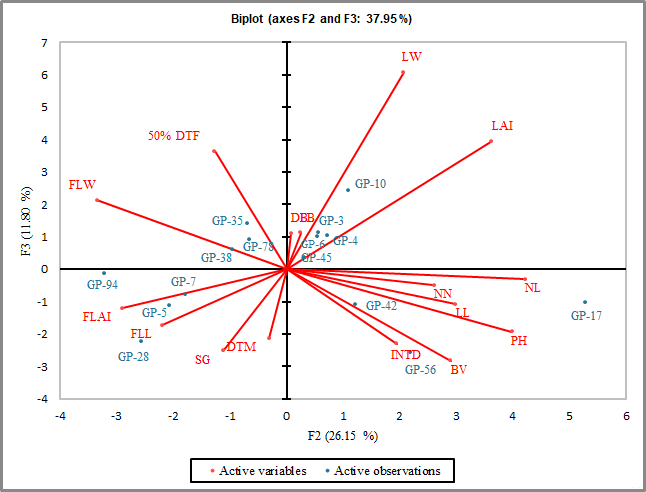
The variables in the biplot were shown to be superimposed as vectors, with the length of each vector representing the proportion of variation in that variable. There was less resemblance between the genotypes that were plotted away from the origin and the genotypes that were plotted closer to the center. 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 1.1). On the other hand, LW trait represented less variability as compared to other traits. Quantitative characters like PH, NL, LAI, NN, BV, FB, LL, LW, NN, and DB and BV were allocated at positive & positive coordinates in biplot analysis in F1 and F2 axes (Figure 1.1). Traits DTM, DTF, FLW, SG, FLL, and FLAI were presented at the fourth positive-negative coordinate in F1 and F2 (Figure 1.1). Only one trait INTD was present at the first negative-positive coordinate in biplot analysis in F1 and F2 axes (Figure 1.1).



### 

### Figure 1.1: Biplot analysis related to different quantitative traits in sorghum for F1 and F2 axes.

The traits like LW and LAI were well represented and exhibited high variability in F2 and F3 axes (Figure 1.2). 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 1.2). Traits DTF and FLW were represented at first negative-positive coordinates in F2 and F3 axes (Figure 1.2). Traits FLAI, FLL, SG, and DTM were allocated at a negative-negative coordinate in biplot analysis in F2 and F3 axes (Figure 1.2). 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 1.2).



### 

### Figure 1.2: Biplot analysis related to different quantitative traits in sorghum for F2 and F3 axes.

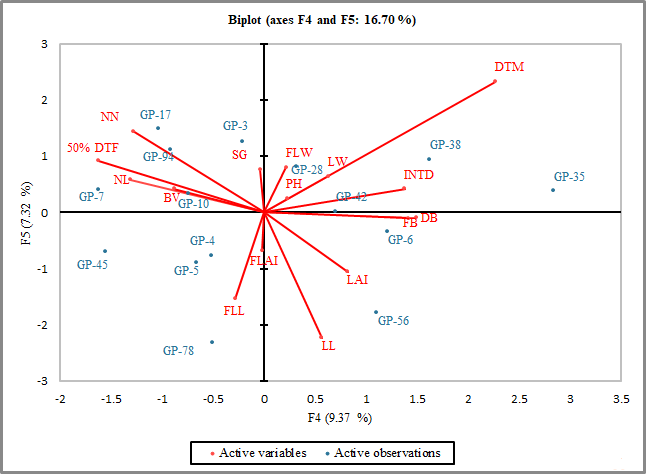
The characteristics such as LAI, LW, FB, and DB were all well recorded and demonstrated a significant degree of variability in both F3 and F4 axes (Figure 1.3). On the other hand, trait FLW presented less variability as related to other traits. Quantitative characters like LAI, LW, FLW, FB and DB, and BV were allocated at a positive-positive coordinate in biplot analysis in F3 and F4 axes (Figure 1.3). Traits DTM, INTD, LL, and PH were represented at a first negative-positive coordinate in F3 and F4 axes (Figure 1.3). 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 1.3). Only DTF trait was represented at the fourth positive-negative coordinate in F3 and F4 axes (Figure 1.3).

### 

### 

### Figure 1.3: Biplot analysis related to different quantitative traits in sorghum for F3 and F4.

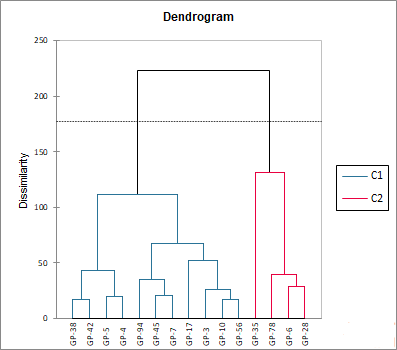
The trait like DTM was well represented and exhibited high variability in F4 and F5 axes (Figure 1.4). There was less variation in the qualities like PH, INTD, LW, and FLW, as compared to other traits. Quantitative characters like DTM, INTD, LW, FLW, and PH were allocated at a positive-positive coordinate in biplot analysis in F4 and F5 axes (Figure 1.4). Traits like SG, BV, NL, DTF, and NN were represented at the first negative-positive coordinate in F4 and F5 axes (Figure 1.4). On the other hand, traits FLAI and FLL were allocated at a negative-negative coordinate in biplot analysis in F4 and F5 axes (Figure 1.4). Traits like DB. FB, LL, and LAI were represented at the fourth positive-negative coordinate in F4 and F5 axes (Figure 1.4).



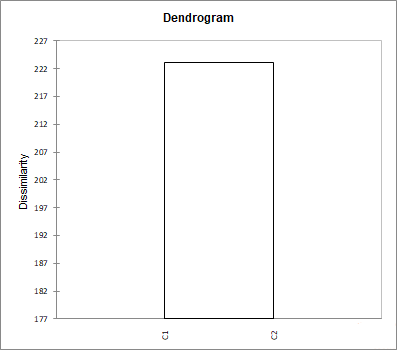
### Figure 1.4: Biplot analysis relating to several quantitative characteristics of sorghum for F4 and F5 axes.

## Cluster analysis

Sorghum genotypes were analyzed using UPGM, which was generated in 5 different morphotypes for one year (Figure 1.5). Cluster analysis revealed that the main cluster was divided into five sub-classes and the clusters C1 and C2 comprised of 11 and 4 genotypes, which had 4 and 1 morphotypes respectively (Figure 1.5). The centroids of the class are described and characterized in (Table 1.6). The genotypes (GP-7) and (GP-6) were screened as central centroids of class III and V, respectively (Table 1.6).



**Figure 1.5: Classification of 15 sorghum genotypes into 5 morphotypes using UPGMA cluster analysis.**

****

**Figure 1.6: Classification of 15 different sorghum accessions from Pakistan based on their genetic similarity as determined by cladogenesis studies.**

## 

**Figure. 1.7: Comparison of 15 sorghum genotypes for fresh biomass and 50% days to flowering.**

## 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 1.5). Stem thickness revealed a lower positive phenotypic correlation with the majority of the traits except for FLL, FLW, FLAI and DTF (negatively correlated). NN showed a high positive phenotypic correlation with plant height (PH) and a lower positive correlation with SG. NN is highly significant and positively correlated with NL. INTD is positively correlated with PH and highly significant and negatively correlated with SG and NL, and LL exhibited a highly significant and positive correlation with NL, SG, NN, and INTD. LW showed a highly significant and positive correlationwith 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 1.5). 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 1.5). 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 1.5).

### Table 1.5: Correlation matrix for 16 quantitative traits of sorghum.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variables** | **PH** | **SG** | **NL** | **NN** | **INTD** | **LL** | **LW** | **LAI** | **FLL** | **FLW** | **FLAI** | **FB** | **50% DTF** | **DB** | **DTM** |
| **PH** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **SG** | 0.091\*\* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **NL** | **0.775\*\*** | 0.060\*\* |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **NN** | 0.395\*\* | 0.264\*\* | **0.783\*\*** |  |  |  |  |  |  |  |  |  |  |  |  |
| **INTD** | 0.386\*\* | -0.271 | -0.009 | -0.229 |  |  |  |  |  |  |  |  |  |  |  |
| **LL** | 0.509\*\* | 0.099 | 0.489\*\* | 0.238\*\* | 0.067\*\* |  |  |  |  |  |  |  |  |  |  |
| **LW** | 0.190\*\* | -0.312 | 0.249\*\* | 0.083\* | 0.073\*\* | -0.058 |  |  |  |  |  |  |  |  |  |
| **LAI** | 0.485\*\* | -0.159 | **0.537\*\*** | 0.245\*\* | 0.069\*\* | **0.647\*\*** | **0.722\*\*** |  |  |  |  |  |  |  |  |
| **FLL** | -0.102 | **0.593\*\*** | -0.136 | -0.001 | -0.509 | 0.099\*\* | -0.334 | -0.179 |  |  |  |  |  |  |  |
| **FLW** | -0.257 | 0.276\*\* | -0.388 | -0.281 | -0.365 | -0.401 | -0.014 | -0.289 | 0.371\*\* |  |  |  |  |  |  |
| **FLAI** | -0.116 | **0.549\*\*** | -0.243 | -0.044 | -0.473 | -0.109 | -0.314 | -0.312 | **0.906\*\*** | **0.644\*\*** |  |  |  |  |  |
| **FB** | 0.243\*\* | 0.486\*\* | 0.188\*\* | 0.355\*\* | -0.386 | 0.277\*\* | 0.212\*\* | 0.364\*\* | 0.512\*\* | 0.304\*\* | **0.541\*\*** |  |  |  |  |
| **50%DTF** | -0.148 | 0.434\*\* | 0.143\*\* | 0.245\*\* | **-0.582** | -0.124 | 0.225\*\* | 0.103\*\* | 0.313\*\* | **0.573\*\*** | 0.387\*\* | 0.222\*\* |  |  |  |
| **DB** | 0.227\*\* | 0.505\*\* | 0.174\* | 0.346\*\* | -0.408 | 0.261\*\* | 0.196\*\* | 0.340\*\* | **0.544\*\*** | 0.336\*\* | **0.578\*\*** | **0.999\*\*** | 0.252\*\* |  |  |
| **DTM** | 0.053\* | 0.297\*\* | -0.110 | 0.060\* | 0.139\*\* | -0.034 | -0.114 | -0.115 | -0.055 | 0.096 | 0.041 | 0.281\*\* | -0.136 | 0.272\*\* |  |
| **BV** | **0.819\*\*** | 0.302\*\* | **0.781\*\*** | **0.533\*\*** | 0.040 | 0.298\*\* | 0.037 | 0.233\*\* | 0.264\*\* | -0.212 | 0.157\*\* | 0.280\*\* | 0.024 | 0.280\*\* | 0.000 |

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

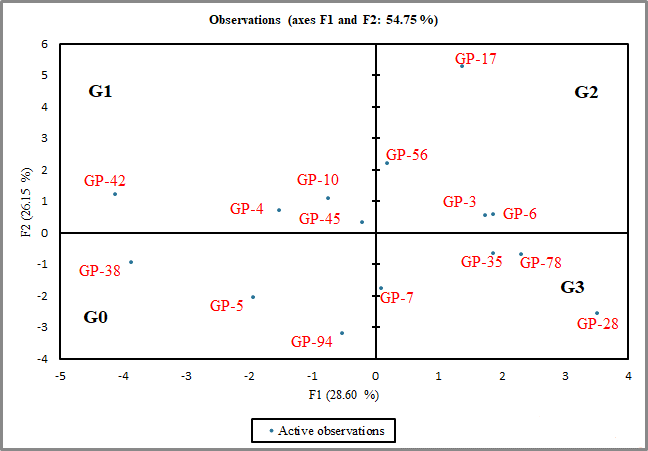
**PH** - Plant height; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf width; **LL** - leaf length; **LW** - Leaf width; **LAI** - Leaf area index; **FLL** - Flag leaf length; **FLW** - Flag leaf width; **FLAI** - Flag leaf area index;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

**Table 1.6:** **Characterization of III and V class centroids in sorghum genotypes.**

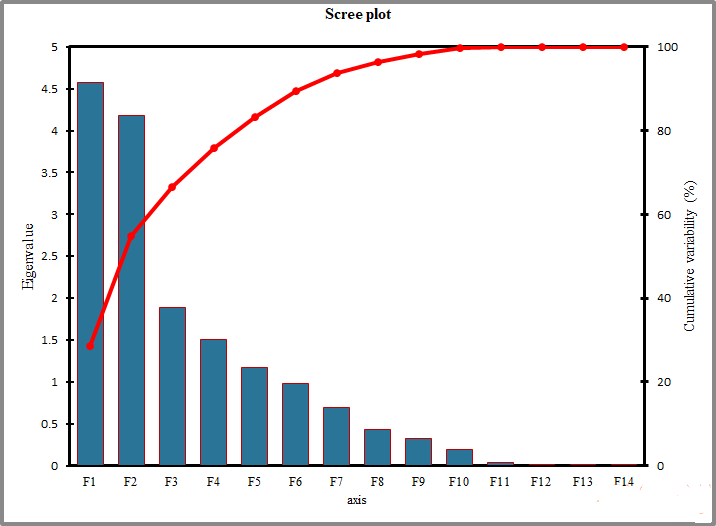
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Cluster** | **PH** | **SG** | **NL** | **NN** | **INTD** | **LL** | **LW** | **LAI** | **FLL** | **FLW** | **FLAI** | **FB** | **50%**  **DTF** | **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 |

## 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) (Fig 1.8). 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 1.8). Three GP-38, GP-5, and GP-94 genotypes were grouped in the G0 (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 1.8). 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 could be considered as sensitive to biotic and abiotic stresses. The G1 including GP- 42, GP-4, and GP-10 were different promising genotypes with better biomass characteristics while in G2, GP-17, and GP-56 and in G3 including GP-28, GP-7, and GP-78 genotypes (Figure 1.8).



**Figure 1.8: PCA grouping of 15 Pakistan sorghum types using discriminatory quantitative traits.**



**Figure 1.9: Scree plot analysis for 15 genotypes.**

**2.4. Identification of desirable genotypes for the sorghum improvement program**

Two genotypes were identified for the sorghum improvement program based on morphological characterization, considering the economic preference characteristics (fresh biomass and days to flowering) of Pakistani farmers. In principle component analysis, the screened 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 highest 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 added 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 (Rakshit *et al.* 2012; Al-Naggar *et al.* 2020). The multivariate analysis tools were used to evaluate the Pakistani sorghum germplasm in the current study. As previously reported by (Jadhav *et al.* 2011; 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 (Adugna 2014). We believe Pakistan's various sorghum germplasm separate the Ethiopian ancestor. Different genotypes may be beneficial for variety development (Jain *et al.,* 2010).

It is easier to discover morphological differences between and across germplasms that have several variates statistical tools (Shrestha 2013; Islam *et al.* 2018). 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 contributor having a share of 26.15%, 11.80%, 9.37% and 7.32% variation in the total variability respectively. Genotypes with high phenotypic variety on the first axis produce more biomass. The third axis can contribute to the biofuel breeding strategy (Maji and Shaibu 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 germplasm in the coordinates (Rakshit and Patil 2013; Azameti *et al.* 2020). We find a positive correlation between several traits, including plant height, days to 50% flowering, and Brix value, these linked traits are essential for the selection of high biofuel genotypes. A similar correlation trend was identified in earlier research (Abubakar and Bubuche 2013; Jain and Patel 2013).

In the current study, 5 morphotypes of sorghum genotypes were yielded with help of UPGMA analysis. To increase the variety of a breeding program, two elite genotypes (GP-7 and GP-6) that belonged to separate groups or morphotypes can be crossed (Jain and Patel 2014; Arshad *et al.* 2017). For three generations, these screened genotypes can be subjected to mass selection to increase homozygosity of the desirable quantitative traits (Rakshit and Patil 2013; Alahmad *et al.* 2018). Due to their better adaptability to changing climate, fresh biomass, and earliness these genotypes can be directly grown to increase sorghum production. In addition, crossing experiments can be designed involving these early maturing genotypes (GP-10, GP-94, GP-17, GP-56, GP-45, GP-78, GP-94, and GP-5) and the two with high fresh biomass (GP-7 and GP-6) to get high biomass and early maturing varieties, as earlier suggested by Dissou-Aminon *et al.* (2015).

* 1. **Conclusion**

Current study estimates largest biggest collection of Pakistan sorghum genotypes so far, for phenotypic diversity assessment. The best performing genotypes discovered related to biomass and biofuels discovered in this study which will improve the present sorghum breeding strategy. Principal component analysis can important to diverse set of genotypes for the association mapping. The cluster analysis extracted class nominators from the five classes, which can provide the novel material with natural resistance against abiotic and biotic stress. Two genotypes GP-6 and GP-7 can be used for direct cultivation in multi-location trials.

**Reference:**

Abubakar L and TS Bubuche (2013). Correlation analysis of some agronomic traits for biomass improvement in sorghum (*Sorghum bicolor L*. Moench) genotypes in north-western nigeria. Afr. J. Agric. Res 8: 3750-3756

Adugna, A (2014). Analysis of in situ diversity and population structure in Ethiopian cultivated *Sorghum bicolor* (*L.)* landraces using phenotypic traits and ssr markers. J. Sprin. Plus 3:212-238

Agrama HA and k Tanistry (2003). Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. Afr. J. Biotechnol 2:334-340

Alahmad S, E Dinglasan, KM Leung, A Riaz, N Derbal, KP Voss-Fels and T Hickey (2018). Speed breeding for multiple quantitative traits in durum wheat. J. Physiol. Mol. Biol 14-1-15

Al-Naggar AM, RR Shafik, RM Musa (2020). Genetic diversity based on morphological traits of 19 maize genotypes using principal component analysis and *GT* biplot. J. Annu. Res. Rev. Biol 10:68-85

Arshad SF, B Sadia, FG Awan and MJ Jaskan (2017). Estimation of genetic divergenceamong sorghum germplasm of Pakistan through multivariate tools. Int. J. Agric. Biol 19:1099-1106

Arshad SF, B. Sadia, FS Awan and MJ Jaskani (2017). Estimation of genetic divergence among sorghum germplasm of Pakistan through multivariate tools. Int. J. Agric. Biol. Eng. 19:1099-1106

Azameti MK, B Vishalakshi, B Umakanth, M Balram, M Srinivas Prasad and MS Madhav (2020). Molecular characterization of popular rice (*Oryza sativa L*.) varieties of India and association analysis for blast resistance. J. Afr. Crop. Sci 8:2225-2236

Dossou-Aminon I, LY Loko, A Adjatin, EE Ewedje, A Dansi, S Rakshit, and K Akpagana (2015). Genetic divergence in northern benin sorghum (*Sorghum bicolor L*. Moench) landraces as revealed by agromorphological traits and selection of candidate genotypes. J. Sci. World 13:17-27

Emendack Y, J Sanchez, C Hayes, M Nesbitt, H Laza and J Burke (2021). Seed-to-seedearly-season cold resiliency in sorghum. J. Front. Plant Sci 15:15-25

Fracasso A, A Perego and S maducci (2017). Characterisation of ten commercial sorghum genotypes grown under water-limited conditions for bioenergy production in Mediterranean environment. J. Proc. Natl Acad. Sci 23:45-67

Hair Jr, RE Anderson, RL Tatham and WC Black (1995). Multivariate data analysis: With readings. Prentice Hall College Division, USA. J. Adv. Life Sci 1:58-59

Islam MZ, M Khalequzzaman, MK Bashar, NA Ivy, MA Mian, BR Pittendrigh and MP Ali (2018). Variability assessment of aromatic rice germplasm by pheno-genomic traits and population structure analysis. J. Front. Plant Sci 8:67-74

Jadhav AR., KN Chaudhari, RT Desai, UR Ghodke and PP Patil (2011). Variability for yield and quality traits in (single-cut) forage sorghum (*Sorghum bicolor L*. Moench). Ind. J. Plant Genet. Res 24:205-210

Jafarzadegan M, AB Safi-Esfahani, Z Beheshti (2019). Combining hierarchical clustering approaches using the PCA method. J. Adv. Life Sci 1:58-59

Jain SK and PR Patel (2014). Characters association and path analysis in sorghum (*Sorghum bicolor L.* Moench) f1s and their parents. J. Ann. Plant Soil Res 16:107-110

Jain SK, M Elangovan and NV Patel (2010). Correlation and path coefficient analysis for agronomical traits in forage sorghum (*Sorghum bicolor L*. Moench). Ind. J. Plant Genet. Res 23:15-18

Johnson RA and DW Wichern (1992). Applied multivariate statistical analysis. J. Sci. World 10:10-19

Kasoma C, H Shimelis, MD Laing, AI Shayano wako and I Mathew (2021). Revealing the genetic diversity of maize (*Zeamays L*.) populations by phenotypic traits and DArTseq markers for variable resistance to fall armyworm. Int. J. Agric. Biol 68:243-259

Maji AT and AA Shaibu (2012). Application of principal component analysis for rice germplasm characterization and evaluation. J. Plant Breed. Crop Sci 4:87-93

Meulman, J (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. Psychometrika. J. Front. Plant Sci 57:539-565

Muhammad A, H Hao, Y Xue, A Alam, S Bai, W Hu and L Wang (2020). Survey of wheat straw stem characteristics for enhanced resistance to lodging. J. Front. Plant Sci 5:2469-2484

Nawaz NA Razzaq, Z Ali, G Sarwar and MY Ousaf (2004). Performance of different at (*Avenasativa*L.) Varieties under the agro-climatic conditions of Bahawalpur, Pakistan.Int. J. Agric. Biol 6:624-626

Pardo J and R VanBuren (2021). Evolutionary innovations driving abiotic stress tolerance in C4 grasses and cereals. J. Plant Biochem. Biotechnol 33:391-401

Rakshit S and JV Patil (2013). Basics of sorghum breeding and aicsip data management. directorate of sorghum research (DSR), India. J. Plant Genet Res 38:83-88

Raza MA, Feng, LY Der Werf, W Iqbal, N Khalid, HB Chen, YK and W Yang (2019). Maize leaf-removal: A new agronomic approach to increase dry matter, flower number and seed-yield of soybean in maize soybean relay intercropping system. J. Amer. Stat. Assoc 9:1-13

Raza MK, MJ Jaskani, SA Naqvi and FS Awan (2020). Exploitation of phenotypic diversity in male accessions of date palm (*Phoenix dactylifera*) and its use in germplasm conservation. Int. J. Agric. Biol 24:133-144

Shrestha J (2013). Agro-morphological characterization of maize inbred lines. J. Agric. Res 2:209-211

Videla ME, J Iglesias and C Bruno (2021). Relative performance of cluster algorithms and validation indices in maize genome-wide structure patterns. J. Euphytica 10:1-19

Ward Jr, JH (1963). Hierarchical grouping to optimize an objective function. J. Amer. Stat. Assoc 58:236-244