Journal of Agricultural Research Advances
ISSN: 2582-7227

Research Article
Open Access

Cluster and principal component analysis of advanced tef [*Eragrostis tef (Zucc.)* Trotter] breeding lines

Girma Ashe[1*], Bulti Tesso[2] and Solomon Chanyalew[1]

[1]Ethiopian Institute of Agricultural Research, Debre Zeit Agricultural Research Center, Addis Ababa, Ethiopia

[2]Haramaya University, Haramaya, Ethiopia

Corresponding author: girmaashe430@gmail.com

Received on: 20/05/2023 Accepted on: 30/08/2023 Published on: 05/09/2023

ABSTRACT

Aim: The aim of this study was to group the breeding lines as their similarities and assess the magnitude of genetic distances among them; then identify the contribution of individual traits to total variations.

Materials and Methods: A total of 49 inbreed lines were evaluated for 17 traits using simple lattice design at Bishoftu and Akaki in 2022 main cropping season. Cluster analysis based on Unweighted Pair Group Method with Arithmetic Means (UPGMA) from Euclidean distances matrix grouped in to five clusters of the 49 tef genotypes.

Results: The members of clusters II and III had mean value greater than overall mean for grain yield. The highest inter-cluster distance was observed between cluster III and IV, while the lowest inter-cluster distances was observed between cluster I and II. Principal components analysis revealed that four principal components with Eigen-values greater than unity accounted for 75% of the variation in tef genotypes.

Conclusion: It was concluded that the genotypes which had larger genetic distance indicate wide genetic divergence, therefore, if such lines are crossed, high variability and better transgressive sergeants may be developed.

Keywords: Breeding Lines, Cluster, Euclidean distance, Genetic distance, Principal component.

How to cite this article: Girma Ashe, Bulti Tesso and Solomon Chanyalew (2023). Cluster and principal component analysis of advanced tef [*Eragrostis tef (Zucc.)* Trotter] breeding lines. J. Agri. Res. Adv., 05(03): 09-16.

Introduction

Tef [*Eragrostis tef (Zucc.)* Trotter], which is indigenous to Ethiopia, is thought to have been domesticated between 4000 and 1000 BC (Vavilov, 1951). Tef is also grown in Eritrea, as well as more recently in US, Netherlands and Israel (Abraham, 2015). It is most significant staple cereal crop in terms of production and consumption and thrives extensively in several climatic and soil environments (Neela and Solomon, 2018). Out of 17.68 million smallholder farmers' households in Ethiopia, 7.2 million grow the crop (CSA, 2020).

Tef is currently drawing attention from all over the world for its nutritional benefits because it is high in nutrients and gluten-free. According to Piccinin (2002), it has 11% protein, 80% complex carbs, and 3% fat. Tef is one of least well-known grains in world, yet it is one of most nutrient grains and *injera* has excellent flavour, aroma, texture and durable quality. It is known for its nutrient quality and 99% flour return upon milling, compared to 60-80%from wheat (Yoseph *et al*., 2020). The grains can also be processed into flour, which is used to make porridge and alcoholic drinks like the local beer "*tela*" and the stronger liquor "*katikala*".

Tef was grown on 3.10 million ha of land in Ethiopia by about 7.15 million smallholder farmers in 2019/20, account for about 24.1% of the country's grain cultivated area. Tef is the most significant economic crop in Ethiopia, where it is grown by 40% of smallholder farmer households and accounts for 17% of all grain production (CSA, 2020). Oromia and Amhara are the two major regions and collectively, the two regions account for 87.2% and 85.3% of the production of tef with an overall average productivity of 1.85 t/ha (CSA, 2020).

The term "cluster analysis" refers to a collection of multivariate approaches whose main objective is to mathematically group together individuals or objects with similar descriptions based on the traits they share (Hair et al., 1995). When individuals are plotted geometrically, those inside one cluster will be closer together and those in other clusters will be farther apart

(Hair et al., 1995). The genetic variability found in conserved germplasm accessions, interspecific and intraspecific hybridization, mutants, and landraces was primarily used to improve other crops. This was made possible by the application of breeding strategies that make use of multivariate statistical analysis (i.e., cluster, distance) as a tool to evaluate the genetic diversity. The expression of sub-traits (yield components) to yield and ease of selection depend on genetic diversity; as a result, it facilitates the effective and efficient utilization of the gene pool in the population. It appears to be useful to classify the existing tef germplasm and recombinant inbred lines into genetically distinct groups.

Selection of genetically divergent parents for hybridization in a transgressive breeding program depends on the proper classification of breeding materials (Sharma, 1998). The varieties must be genetically divergent in order to create a successful hybridization program, particularly for quantitative traits that affect yield (Singh, 1983). As a result, crossings between populations with the highest genetic divergent would be more responsive to improvement since their offspring are more likely to exhibit optimal recombination and segregation following hybridization. Therefore, in order to give new opportunities for recombination and selection, genetic diversity must be periodically introduced into the population in each breeding effort (Welsh, 1981).

Additionally, understanding the naturally occurring diversity in a population helps in the identification of several tef genotypes with regard to high grain yield, tolerance or resistance to low moisture, lodging resistance, earliness, and desired grain quality (Kebebew et al., 2013). So in order to effectively utilise the population's gene pool for breeding purposes in the future, it is necessary to assess the genotypes that are currently available for genetic divergence. Therefore, the current study was conducted to group the lines as their similarities and assess the magnitude of genetic distance among them and then identify the contribution of individual traits for total variations.

Materials and Methods

A field study was conducted during the 2022 mean cropping season at two locations (Debre Zeit Agricultural Research Center (DZARC) main station (Bishoftu) and Akaki sub-station). DZARC is found at (8° 44′ N, 38° 58′ E, and 1900 m.a.s.l), whereas Akaki is found at (8º 53' N, 38° 58′ E, and 2400 m.a.s.l) latitude, longitude, and altitude, respectively. The two locations are characterized by a moist tropical climate and experience a long rainy season extending from June to September. Bishoftu receives a mean annual rainfall of 832 mm during the main growing season, with maximum and minimum mean annual temperatures of 24.3 °C and 8.9 °C, respectively. In contrast, Akaki often receives annual total rainfall of 1254 mm with maximum and minimum mean annual temperatures of 30 °C and 10 °C, respectively. The experimental field at both locations is characterized by heavy black soil (vertisoil) with a very high moisture retention capacity.

Experimental Plant Materials

The experimental tef plant materials were obtained from Debre Zeit Agricultural Research Center of National Tef Breeding Program. Totally, forty-nine genotypes (forty-eight advance line and one standard check (Dagim)) were used in the experiment.

Table 1. List and description of tef genotypes used for the study.

No.	Pedigree (Genotype)	No.	Pedigree (Genotype)
1.	DZ- Cr-387xRosea (RIL 162)	26.	DZ-01-1681 x Alba (RIL 147)
2.	DZ- Cr-387xRosea (RIL 14)	27.	DZ-01-1681x Alba (RIL 142)
3.	DZ- Cr-387xRosea (RIL 106)	28.	DZ-01-1681x Alba (RIL 144)
4.	DZ- Cr-387xRosea (RIL 196)	29.	DZ-01-1681 x Alba (RIL 31)
5.	DZ- Cr-387xRosea (RIL 173)	30.	DZ-01-1681 x Alba (RIL 87)
6.	DZ- Cr-387xRosea (RIL 6)	31.	DZ-01-1681 x Alba (RIL 175)
7.	DZ- Cr-387xRosea (RIL 132)	32.	DZ-01-1681 x Alba (RIL 103)
8.	DZ- Cr-387xRosea (RIL 92)	33.	DZ-01-1681 x Alba (RIL 76)
9.	DZ- Cr-387xRosea (RIL 96)	34.	DZ-01-1681 x Alba (RIL 121)
10.	DZ- Cr-387xRosea (RIL 117)	35.	DZ-01-1681 x Alba (RIL 32)
11.	DZ- Cr-387xRosea (RIL 138)	36.	DZ-01-1681 x Alba (RIL 78)
12.	DZ- Cr-387xRosea (RIL 163)	37.	DZ-01-1681 x Alba (RIL 47)
13.	DZ- Cr-87xRosea (RIL 7)	38.	DZ-01-1681 x Alba (RIL 70)
14.	DZ- Cr-387xRosea (RIL 58)	39.	DZ-01-1681 x Alba (RIL 97)
15.	DZ- Cr-387xRosea (RIL 107)	40.	DZ-01-1681 x Alba (RIL 116)
16.	DZ- Cr-387xRosea (RIL 53)	41.	DZ-01-1681 x Alba (RIL 46)
17.	DZ- Cr-387xRosea (RIL 122)	42.	DZ-01-1681 x Alba (RIL 30)
18.	DZ- Cr-387xRosea (RIL 119)	43.	DZ-01-1681 x Alba (RIL 15)
19.	DZ- Cr-387xRosea (RIL 1)	44.	DZ-01-1681 x Alba (RIL 100)
20.	DZ- Cr-387xRosea (RIL 98)	45.	DZ-01-1681 x Alba (RIL 134)
21.	DZ- Cr-387xRosea (RIL 157)	46.	DZ-01-1681 x Alba (RIL 185)
22.	DZ- Cr-387xRosea (RIL 155)	47.	DZ-01-1681 x Alba (RIL 2)
23.	DZ- Cr-387xRosea (RIL 166)	48.	DZ-01-1681 x Alba (RIL 48)
24.	DZ-Cr-387xRosea (RIL 91)	49.	Dagim (DZ-Cr-438 RIL91)
25.	DZ-01-1681 x Alba (RIL 120)		

Experimental Design, Layout and Management

The experiment was laid out in 7x7 simple lattice designs with two replications. Each experimental plot was $2m^2$ (1mx2m) and consist of five rows spaced 20 cm apart. Incomplete block distances and plot distances within incomplete block distances were 1.5 m and 1 m, respectively. Within each replication, the genotypes were assigned to plots at random. All others crops management practices and recommendations were applied uniformly to all genotypes as recommended for the crop.

Data Collection and Analysis

Days to 50% heading, days to 90% physiological maturity, grain filling period, plant height, panicle length, peduncle length, culm length, number of spikelets per panicle, number of primary panicle branches per main shoot, number of florets per spikelet, number of total tillers per plant, number of fertile tillers per plant, lodging index, above-ground biomass per plot, grain yield per plot, harvest index and thousand seed weight data were collected as recommended for the tef and subjected to analysis using appropriate software.

Cluster and Genetic Divergence Analyses

For cluster analysis, mean records for all traits were pre-standardized to mean zero and variance unity in order to remove bias resulting from differences in measuring scales (Manly, 1986). The cluster analysis was performed based on Unweighted Pair Group Method with Arithmetic Means (UPGMA) clustering method from Euclidean distance matrix following the average linkage method by SAS software (v.9.2). Genetic distance of the tef genotypes were estimated using Euclidean distance (ED) calculated from quantitative traits after standardization (subtracting the mean value and dividing it by the standard deviation) as established by Mohammadi and Prasanna (2003) as follows;

$$ED_{jk} = \sqrt{\sum_{i=j}^{n} (Xji - Xki)^2}$$

Where; EDjk = distance between genotypes j and k; Xij and Xik = phenotype traits values of the ith character for genotypes j and k, respectively; and n = number of phenotype traits used to calculate the distance. The calculated average distance (ED) was used to estimate which genotype(s) are closest or distant to others.

Principal Components Analysis

For principal component analysis, the pre-standardized trait mean data of the test tef genotypes was utilized to determine the key traits that accounted for the majority of the gross observed variability among the genotypes. The principal component variables are defined as linear combination of the original variables *Xl …, Xk…, Xm*. For the equations below, coefficients are provided by the table of retrieved eigenvectors.

Yk= Ck1X1 + Ck2X2 + … CkmXm

Where; Yk is the kth principal component k and C's are the coefficients in table.

According to Holland (2008) suggested standard criteria that permit to ignore components whose variance explained is less than 1 when a correlation matrix is used for determining number of PCs should be investigated was employed. According to Sharma (1998), principal component analysis highlights the significance of the important of traits that contribute the most to the overall variation along each differentiation axis. Principal component analysis was computed by using SAS computers software. We considered principal components with eigenvalues greater than unity and, following Johnson and Wichern (1988), we took the eigenvectors with values greater than half divided by the standard deviation or the square root of the corresponding eigenvalue to identify the traits that contributed to each PC.

Results and Discussion

Clustering of Genotypes

The cluster analysis based on Unweighted Pair Group Method with Arithmetic Means (UPGMA) clustering method using Euclidean distances matrix grouped the 49 tef genotypes into five major clusters consisting of 1 to 15 genotypes (Table 1 and 2).Cluster II and IVwere the largest cluster consisted of fifteen genotypes (30.61%) of the total genotypes with the standard check (Dagim) grouped in Cluster IV.

Cluster I comprised seven genotypes (14.29%) of total genotypes, whereas cluster III consisted one genotype (2%). Cluster V consisted of eleven genotypes (22.45%) of the total genotypes (Table 1 and 2). Such genetically divergent tef genotypes indicated that crossing between genotypes of these clusters might provide desirable recombinants and high yielding segregants. Similar works had been reported previously by Getahun (2021) who grouped 49 tef genotypes into four clusters. Habte *et al*. (2015); Habte *et al*. (2017) grouped 26

and 36 tef genotypes into six clusters. Worku and Kebebew (2021) also classified 49 tef genotypes in to twelve distinct clusters under drought stress.

The findings indicated that, in majority of cases, genotypes resulting from crosses of same parents grouped together, it may be due to an interchange of genetic material between parents. The different genotypes grouped within a given clusters were assumed to be more closely related in term of studied traits than those genotypes grouped into different clusters. The current cluster analysis revealed that there is variability even among the lines resulting from the crosses of the same parents.

Cluster Mean Analysis

The mean values of 17 quantitative characters in each cluster were presented (Table 3). Cluster I had mean values greater than overall mean for days to heading, days to physiological maturity, grain filling period, peduncle length, number of florets per spikelet, and above ground biomass, but for rest of traits it had mean values lower than overall mean.Cluster II had mean values greater than overall mean for number of florets per spikelets, number of total tillers per plant, number of fertile tillers per plant, lodging index, above-ground biomass and grain yield. Cluster III had mean values greater than overall mean for grain filling period, plant height, culm length, number of florets per spikelets, lodging index, grain yield and harvest index (Table 3).

Cluster IV was characterized by having greater mean values for days to heading, days to physiological maturity, grain filling period, plant height, panicle length, culm length, peduncle length, number of spikelets per panicle, number of primary panicle branches per main shoot and above ground biomass than overall mean. Cluster V had mean values greater than overall mean for days to heading, number of spikelets per panicle, number of primary panicle branches per main shoot, number of total tillers per plant, number of fertile tillers per plant, lodging index and above ground biomass. The members of clusters I, IV and V had mean value lower than overall mean for grain yield (Table 3). Desheva and Cholakov (2014) reported one cluster among three clusters suitable for source of hybridization. Salman *et al*. (2014) identified one cluster among six clusters.

Table 2. Distribution of 49 tef genotypes into five different clusters based on UPGMA of means of 17 traits over two locations.

Cluster	Number and % of genotypes	List of Genotypes
I	7(14.29%)	G1, G2, G11, G13, G20, G43, G46
II	15 (30.61%)	G3, G4, G5, G6, G7, G9, G10, G12, G15, G16, G18, G19, G22, G24, G39
III	1(2%)	G8
IV	15(30.61%)	G14, G17, G21, G29, G31, G33, G34, G35, G36, G37, G38, G40, G45, G47, G49
V	11 (22.45%)	G23, G25, G26, G27, G28, G30, G32, G41, G42, G44, G48

G=Genotype code used in the experiment

Table3. Cluster mean of 49 genotypes for 17 traits of average over two test locations.

Traits	Cluster means					
	C I	C II	C III	CIV	C V	Overall mean
DTH	51.12	50.47	48	54.8	51.96	51.27
DTM	112.21	110.28	108	115.33	107.5	110.664
GFP	60.86	59.88	60.5	60	55.27	59.302
PH	101.15	98.72	106	110.03	97.86	102.752
PL	39.51	39.26	40.3	45.14	40.13	40.868
CL	65.7	59.46	65.7	64.88	57.73	62.694
PDL	22.56	20.11	19.4	20.66	20.49	20.644
NSPP	562.47	548.03	590.25	710.91	657.23	613.778
NPPBPMS	18.31	18.22	18.5	24.69	21.66	20.276
NFPS	6.29	6.05	6.08	5.42	5.45	5.858
NTTPP	11.5	13.76	11.35	12.44	12.9	12.39
NFTPP	9.77	11.69	9.4	10.51	10.65	10.404
LI	68.93	77.5	74.5	55.25	67.73	68.782
BY	13251.8	13665.8	10750	13645.8	13105.1	12883.7
GY	2598.3	2712.2	2875.5	2559.1	2547.9	2658.6
HI	19.63	19.87	37.57	19.78	19.49	23.268
TSW	0.296	0.298	0.29	0.303	0.264	0.2902

C= cluster, DTH =days to heading, DTM = days to physiological maturity, GFP = grain filling period, PH= Plant height, PL=panicle length, CL= culm length, PDL= peduncle length, NSPP=number of spikelets per panicle, NPPBMS = number of primary panicle branches per main shoot, NFPS =number of florets per spikelet, NTTPP= number of total tillers per plant, NFTPP= number of fertile tillers per plant, LI= lodging index, BY=biomass yield, GY= grain yield, HI = harvest index and TSW= thousand-seed weight.

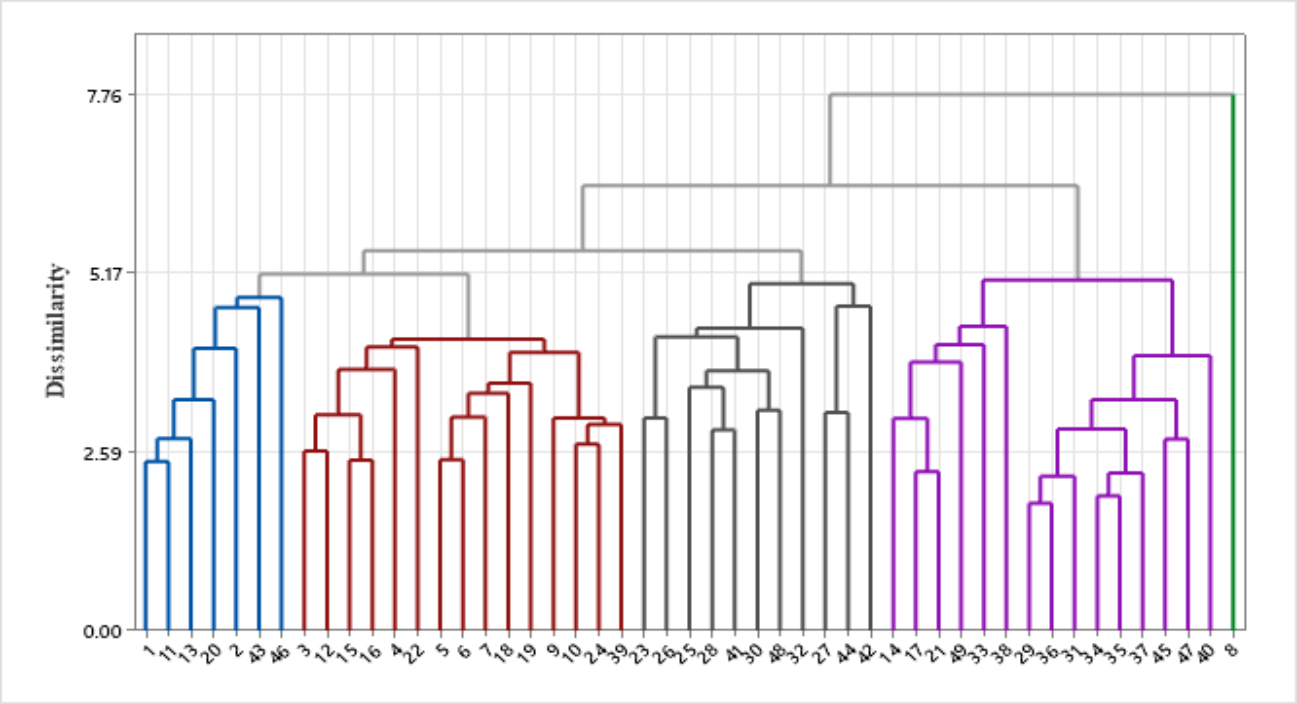

Fig 1. Dendrogram depicting dissimilarity of 49 tef genotypes based on Unweighted Pair Group Method with Arithmetic Means (UPGMA) using Euclidean distances matrix from 17 traits.

Distance among Clusters

The degree of improvement gained through selection and hybridization depends on the range of variability existing among genotypes. The larger the distance between two clusters, the wider the genetic variability between them to be included in hybridization program (Parameshwarappa et al., 2010).The highest intra-cluster distance was observed for cluster V, followed by IV, II and I, while the lowest intra-cluster distanceswere observed for cluster III. The highest inter-cluster distance was observed between cluster III and IV followed by I and III, III and V and II and III, while the lowest inter-cluster distances were observed between cluster I and V, I and IV, II and IV and IV and V (Table 4). Similarly, Habte *et al*. (2017) and Getahun (2021) used D^2 statistics to determine intra- and inter-cluster distances among clusters in tef.Genetic improvement through hybridization and selection depends on the extent of variability among the lines. Therefore the, current findings suggested that the presence of genotypes that are genetically distant from one another is crucial in hybridization programs that aim to combine the best traits to create varieties

Principal Component Analysis (PCA)

Interpretation of the principal components is based on the finding which of the studied variables is most powerfully correlated with each component (i.e. which of these eigenvectors are large in magnitude, the farthest from zero in either positive or negative direction). Holland (2008) stated that when using a correlation matrix, standard criteria allow for the elimination of components with Eigen values less than 1.In the present study, the principal components analysis revealed that four principal components with eigenvalues greater than unity accounted for about 75% of total variation for 17 traits, indicating presence of genetic variability among genotypes studied for the traits considered (Table 5). Similarly Habte*et al*. (2017) reported that about 85% of the total variation among 28 tef varieties evaluated for 16 traits was explained by five principal components. Likewise, Getahun (2021) reported that 77.63% of the total variation among 49 tef varieties evaluated for 15 traits was explained by five principal components.

The first principal component alone explained 36% of the total variation, while PC2, PC3 and PC4 in that order accounted for 18%, 13% and 8% of the gross observed variation among the test tef genotypes. Days to heading, days to physiological maturity, plant height, panicle length, culm length, number of spikelets per panicle and number of primary panicle branches per main shoot showed greater loading

on the first PC. Similarly, days to physiological maturity, grain filling period, plant height, culm length, number of spikelets per panicle, above-ground biomass, grain yield and thousand seed weight contributed most to the second PC; while in the third PC, number of total tillers per plant, number of fertile tillers per plant, above ground biomass and grain yield accounted for much of the observed gross variation. In the fourth PC, peduncle length, number of florets per spikelet, numberof total tillers per plant and number of fertile tillers per plant was the important trait accounted for much of the observed gross variation.

First two PCs accounted for a 54% of total variation indicating much of variability observed originated from traits included in these PCs. Most of important yield and yield attributing traits were present in first two PCs. High PC score for a given genotype in a particular component denotes high values for variables in that particular genotype. There was a sharp decline in contribution from PC1 to PC2 and then from PC2 to PC3 in that order whiles rate of decrease in contribution became lower for remaining PCs. It showed that first few principal components had greatest contribution to overall variation among lines in term of 17 traits considered in this study.

Table 4. Average intra- (bold) and inter- (off) diagonal cluster distance among five clusters of 49 tef genotypes based on average data of 17 traits over two test location.

	C I	C II	C III	C IV	C V
C I	**3.58**	18.39	209.82	44.65	29.26
C II		**3.64**	176.61	47.39	21.06
C III			**0.00**	250.58	195.64
C IV				**3.81**	52.3
C V					**4.01**

Table 5. Eigenvectors, eigenvalues and percentage of total variance explained by the first fourprincipal components (PC) for 17 traits of 49 tef genotypes (average over two locations.

Traits	Eigenvectors			
	PC1	PC2	PC3	PC4
Day to heading (days)	0.857	-0.217	-0.091	-0.013
Days to physiological maturity (days)	0.767	0.461	0.042	0.093
Grain filling period (days)	0.269	0.798	0.010	0.133
Plant height (cm)	0.833	0.388	-0.047	-0.066
Panicle length (cm)	0.882	0.256	0.103	-0.110
Culm length (cm)	0.556	0.496	-0.064	-0.003
Peduncle length (cm)	0.088	0.141	-0.600	0.572
No. of spikelets per panicle	0.743	-0.227	0.207	-0.425
No. of primary panicle branches per main shoot	0.848	-0.285	0.095	-0.097
No. of florets per spikelet	-0.450	0.360	-0.292	0.441
No. of total tillers per plant	-0.266	-0.106	0.851	0.290
No. of fertile tillers per plant	-0.174	-0.046	0.856	0.363
Lodging index (%)	-0.870	0.210	0.114	-0.098
Above ground biomass (kg/ha)	0.138	0.538	0.455	0.058
Grain yield (kg/ha)	-0.527	0.519	0.213	-0.428
Harvest index (%)	-0.580	0.131	-0.222	-0.549
Thousand-seed weight (g)	0.139	0.732	-0.004	-0.014
Eigenvalue	6.46	3.17	2.30	1.51
Differences	3.29	0.87	0.79	0.56
Variance explained (%)	36.00	18.00	13.00	8 .00
Cumulative variance explained (%)	36.00	54.00	66.00	75.00

Conclusions

Genetic characterization and evaluation of indigenous germplasm resources and/or genotypes are very essential towards the development of new tef varieties with traits of interest. Cluster analysis grouped the breeding lines into five clusters based on their similarity. The highest intra-cluster distance was observed for cluster V, while the lowest was for cluster III. The highest inter-cluster distance occurred between clusters III and IV, while the lowest one was between clusters I and II. Principal components analysis showed that about 75 % of the gross variance among lines was laid in PC1 to PC5, and the total variance was loaded largely by traits like days to physiological maturity, grain filling period, plant height, panicle length, culm length, number of fertile tillers per plant, and above ground biomass.The first two PCs accounted for 54% of the total variation, indicating that much of the variability observed originated from traits included in these PCs. The genotypes which had larger genetic distance indicate wide genetic divergence, therefore, if such lines are crossed, high variability and better transgressive sergeants may be developed.

References

Abraham R (2015). Achieving food security in Ethiopia by promoting productivity of future world food tef: A review. Advances in Plants & Agriculture Research, 2(2): 45.

CSA (2020). Central Statistical Agency, Agricultural sample survey 2019/2020 (2013 E, C,). Volume I. Report on area and production of major crops (Private Peasant Holdings, *Meher* season). Statistical Bulletin 578 Addis Ababa, Ethiopia.

Desheva G and Cholakov T (2014). Variability, heritability and genetic progress for some yield components in common winter wheat genotypes (*Triticum aestivum* L.). Genetics and Plant Physiology, *4*(3-4): 191-200.

Getahun Bekana (2021). Cluster and Principal Component Analysis of Semi-Dwarf Tef [*Eragrostis tef* (Zucc.) Trotter] Recombinant Inbred Lines with Emphasis to Lodging. American Journal of Biomedical and Life Sciences. 9(4): 182-189.

Habte Jifar, Kasahun Tesfaye, Kebebew Assefa, Solomon Chanyalew and Zerihun Tadele (2017). Semi-dwarf tef lines for high seed yield and lodging tolerance in Central Ethiopia. African Crop Science Journal, *25*(4): 419-439.

Habte Jifar, Kebebew Assefa and Zerihun Tadele (2015). Grain yield variation and association of major traits in brown-seeded genotypes of tef [*Eragrostis tef* (Zucc.) Trotter]. Agriculture & Food Security, 4(1):1-9.

Hair JR, Anderson RE, Tatham RL and Black WC (1995). Multivariate data analysis with readings. 4th edition, Prentice-Hall, Englewood Cliffs, N.J.

Holland SM (2008). Principal components analysis (PCA). Department of Geology, University of Georgia, Athens, GA, 30602 -250.

Johnson RA and Wichern DW (1988). Applied Multivariate Analysis. 2nd Edition, John Wiley & Sins Inc., New York, USA.

Kebebew Assefa, Merker A and Hailu Tefera (2003). Multivariate analysis of diversity of tef (*Eragrostis tef* (Zucc.)Trotter) germplasm from western and southern Ethiopia. Hereditas, 138 (3): 228-236.

Manly BFJ (1986). Multivariate Statistical Methods: A Primer. Chapman and Hall. London, UK.

Mohammadi SA and Prasanna BM (2003). Analysis of genetic diversity in crop plantssalient statistical tools and considerations. Crop Science, 43(4): 1235-1248.

Neela, S and Solomon Workineh Fanta (2018). Review on structural, nutritional and anti-nutritional composition of tef (*Eragrostis tef*) in comparison with quinoa (*Chenopodium quinoa* Willd.). Cogent Food & Agriculture, 4(1): 1546942.

Parameshwarappa SG, Palakshappa MG and Salimath PM (2010). Analysis of genetic divergence in sesame (*Sesamum indicum* L.). Karnataka Journal of Agricultural Science, 23: 227-230.

Piccinin D (2002). More about Ethiopian food: Tef. University of Washington, USA. http //ethnomed.org/clinical/nutrition/more-about-Ethiopian-food-tef. (Accessed on February, 27.2017).

Salman S, Shah J, Khan J, Rehmat K, Khan U and Khan I (2014). Genetic variability studies in bread wheat (*Triticum aestivum* L.) accessions. Pakistan Journal of Agricultural Research, *27*(1):1-7.

Sharma JR (1998). Statistical and Biometrical Techniques in Plant Breeding. *New Age International Publishers*, New Delhi, 432 pp.

Singh BD (1983). Plant breeding principles and methods. Kalyani Publisher, New Delhi, India. 506 pp.

Vavilov NI (1951). The Origin, Variation, Immunity and Breeding of Cultivated Plants. Translated from the Russian by Starch ester K., Ronald Press, New York, pp 37-38.

Welsh JR (1981). Fundamentals of Plant Genetics and Breeding. John Willey and Sons, Inc., New York, 290 pp.

Worku Kebede and Kebebew Assefa (2021). Cluster analysis of drought tolerant tef genotypes. Journal of Agricultural Research Advance, 03(01): 12-20.

Yoseph Asmelash, Desta Berheand Kwang-Pyo Kim, 2020. Nutritional composition and health benefits of tef (Eragrostis tef (Zucc.) Trotter). Journal of Food Quality, *2020*.

****