
Principal Component Analysis of Early Generation Drought Tolerant Tef Genotypes for Yield-contributing Traits

Worku Kebede^{1,*}, Bulti Tesso²

¹Ethiopian Agricultural Research Institute, Debre Zeit Agricultural Research Center, Debre Zeit, Ethiopia

²School of Plant Sciences, Haramaya University, Dire Dawa, Ethiopia

Email address:

workukebede121@gmail.com (W. Kebede)

*Corresponding author

To cite this article:

Worku Kebede, Bulti Tesso. Principal Component Analysis of Early Generation Drought Tolerant Tef Genotypes for Yield-contributing Traits. *International Journal of Science, Technology and Society*. Vol. 9, No. 3, 2021, pp. 113-118. doi: 10.11648/j.ijsts.20211003.11

Received: March 17, 2021; Accepted: May 25, 2021; Published: June 3, 2021

Abstract: The present investigation was carried out to determine the relationship and genetic variability among 49 tef inbred line using principal component analysis for drought prone areas. To improve tef productivity, farmers need high-yielding and drought tolerant tef cultivars. The objective of this research is to evaluate genetic diversity among drought tolerant tef inbred lines for yield, yield-contributing traits. In this study, Component I had the contribution from the traits *viz.*, days to heading, days to physiological maturity, plant height, panicle length, culm length, number of spikelets per panicle, number of primary panicle branches per main shoot, lodging index, above-ground biomass and harvest index which accounted 40% to the total variability. Grain filling period, number of total tillers per plant, number of fertile tillers per plant, days to mature, peduncle length, number of florets per spikelet and thousand-seed weight has contributed 14% to the total variability in component II. The remaining variability of 13%, 7% and 6% was consolidated in component III, component IV and component V by various traits like days to seedling emergence, culm length, peduncle length, lodging index, above-ground biomass yield, grain yield, harvest index, number of total and fertile tillers per plant. The cumulative variance of 79% of total variation among 18 characters was explained by the first five axes. Thus, the results of principal component analysis revealed, wide genetic variability exists in this drought tolerant tef inbred lines. Drought tolerant traits with high genetic variability are expected to provide high level of gene transfer during breeding programs.

Keywords: Tef, PCA, Drought Tolerant, Inbred Lines and Association

1. Introduction

Tef [*Eragrostis tef* (Zucc.) Trotter] is endemic to Ethiopia and its domestication is estimated to have occurred between 4000 and 1000 BC [26]. Tef is also cultivated in very small quantities in Eritrea and recently in the USA, the Netherlands and Israel [2]. Ethiopia, where tef is the main cereal crop and food shortage is a recurring phenomenon, exerted an export ban on tef which increased interest in growing tef outside Ethiopia.

Currently, the crop is increasingly receiving global attention for its nutritional advantages because it is rich in nutrients and is gluten free. Consumers prefer tef due to its high protein, high mineral content and good quality “*injera*”, a pancake-like soft bread [7] and the absence of gluten [24], which makes it an alternative food for people suffering from

celiac disease. Due to this “life-style” nature of the crop, it has been heralded as a super food or super grain for human being [11, 22]. It contains 11% protein, 80% complex carbohydrates and 3% fat [20].

Tef grows under a wide range of ecological conditions, from sea level up to 3000 meters above sea level (m.a.s.l). Tef has the genetic potential to yield up to 6 t ha⁻¹ [23]. Despite its numerous relative advantages and economic importance, the productivity of tef in Ethiopia is low, amounting to 1.85 t ha⁻¹[6]. The major yield limiting factors to tef production are lack of cultivars tolerant to lodging and drought conditions [13], as well as small seed size. Yield losses are estimated to reach up to 40% during severe moisture stress [18].

To maximize selection benefits, breeding for drought tolerance requires the accumulation of additive genes, a

controlled stress screening environment, and high throughput selection methods [4]. While early research in tef revealed significant genotypic differences in drought tolerance in relation to root depth and osmotic adjustment [17], information on drought tolerance based on grain yield is scarce. Drought, salinity, and heat, as well as climate change, have a significant impact on crop production and food security. It has been proposed that future studies should concentrate on improving resistance or tolerance to these environmental calamities [28].

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables [12]. Principal component analysis is an important breeding tool commonly used by breeders to identify traits that could be used to discriminate crop genotypes [27]. Establishing suitable selection criteria for identifying genotypes with desirable traits is useful in developing improved varieties. In order to plan an effective breeding program, it is critical to analyze trait variability and understand the relationships between traits that contribute to yield [16].

Principal component analysis (PCA) analyzes a data table representing observations described by several dependent variables, which are, in general, inter-correlated. Its goal is to extract the important information from the data table and to express this information as a set of new orthogonal variables called principal components [1, 21] found that four PCs explained 80% of the variation of 13 traits for tef landraces grown under greenhouse conditions. In their studies, grain color, days to maturity, the number of panicles and number of internodes per plant, second culm internode diameter, plant height, shoot biomass, grain yield, and harvest index were well correlated with PC1, which explained 40% of the variation. [15] reported that five PCs explained 71% of the variation of 17 quantitative traits found in 320 tef lines and 35 landraces, evaluated at two locations in central Ethiopia. These authors found that PC1 was correlated with the number of productive tillers per plant, grain yield and the harvest index.

The principal components is used to interpreted based on finding which variables are most strongly correlated with each component. According to [8] Eigenvalues greater than one were only for the first three PCs, which together explained 75% of the observed variation. [21] observed four principal components (PCs), having eigenvalues between 5.16 and 1.12. According to [9] report the first three principal components (PCs) with eigenvalue greater than one contributed for 78.3% of the entire phenotypic variation observed among the 36 tef genotypes. The study is aimed to determine variation of drought tolerant tef genotypes to identify and classify variation for grouping the inbred lines by taking into account several characteristics and relationship

between them with the help of principal component analysis

2. Material and Method

2.1. Experimental Sites, Designs and Experimental Materials

A field study was conducted during the 2017 cropping season at two locations (Melkassa 8° 24' N, 39° 21' E and Alemtena 8° 20' N, 38° 57' E) in the Central Rift Valley of Ethiopia. Both Alemtena and Melkassa are drought prone areas, with Andosols of very light texture that show low water retention capacity [3]. The region experiences poor rainfall distribution (500 mm to 750 mm), coupled with relatively high temperature (15°C to 30°C), which makes the area vulnerable to moisture stress.

For this study, 49 genotypes, including 42 drought tolerant advanced lines (*Dtt*), four parents of the advanced lines, two varieties and a local check were used for this study. *Dtt*₂ (drought tolerant tef 2) and *Dtt*₁₃ (drought tolerant tef 13) were obtained from ethylmethane sulfonate mutagenized populations of Tsedey using the targeted induced local lesions IN genomes (TILLING) method at the Institute of Plant Sciences of the University of Bern through the Tef Improvement Project supported by the Syngenta Foundation for Sustainable Agriculture. These lines are depicted excellent performance under moisture scarcity. The unique morphological difference between the *Dtt* and the original parental tef line Tseday (DZ-Cr-37) is the size and number of stomata. The stomata at the adaxial or upper side of the two *Dtt* lines are smaller both in size and number compared to the original parental tef line [5].

Debre Zeit Agricultural Research Center provided seeds for all genotypes. The experiment was laid out in a 7 x 7 simple lattice design. Each experimental plot was 1 m² (1 m x 1 m) and consisted of five rows spaced 20 cm apart. The distances between both incomplete blocks and plots within incomplete blocks were 1 m, and that between replications was 1.5 m. Seeds were sown at the recommended rate of 15 kg ha⁻¹, amounting to 1.5 g of seed per plot per row. The recommended full doses of blended fertilizer urea (21.74 kg) and NPS (158 kg) per hectare were applied at both locations.

2.2. Data Collection and Analysis

Data were recorded for days to 50% seedling emergence, days to 50% heading, days to 90% physiological maturity, grain filling period, plant height, panicle and, 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 (%), total above-ground biomass, total grain yield, harvest index (%), thousand grain weight. The principal component variables are defined as linear combinations of the original variables $X_1 \dots, X_k \dots, X_m$. The extracted eigenvectors table provides coefficients for Equations [19].

$$Y_k = C_{k1}X_1 + C_{k2}X_2 + \dots + C_{km}X_m$$

Where: Y_k = the k^{th} principal component k ; and C 's = the coefficients

[10] Suggested standard criteria that allow to overlook components whose variance explained is less than one when a correlation matrix is used for determining number of PCs should be investigated was employed. It also indicates that data were standardized to mean zero and variance of one before computing principal component analysis. The principal component based on correlation matrix, was calculated using MINITAB software.

3. Result and Discussion

3.1. General Principal Component Analysis (PCA)

The principal components analysis revealed that five principal components with Eigen-values greater than unity accounted for 79 percent of the gross variability in 18 phenomorphic characters (Table 1). [10] suggested that standard criteria permit to ignore components whose Eigen values are less than 1 when a correlation matrix is used. Similarly [14, 15], reported that about 71-79 percent of the variation in 320 tef germplasm lines was explained by five PCs. Likewise, [25] reported that 76 percent of the total variation among 49 tef varieties evaluated for 23 traits was explained by six PCs. The cumulative variance of 79 percent (Table 1) by the first five axes with Eigen value of > 1.0 indicates that the identified traits within the axes exhibited great influence on the phenotype of germplasm lines.

3.2. Scree Plot of Principal Component Analysis (PCA)

The Scree plot of the PCA (Figure 1) shows that the first five eigenvalues correspond to the whole percentage of the variance in the dataset. The first principal component alone explained 40 percent of the total variation, while PC2, PC3, PC4 and PC5 in that order accounted for 14, 13, 7, and 6 percent of the gross observed variation among the test tef genotypes. The first three PCs, together accounted for a cumulative of 67 percent of the total variation indicating that much of the variability among the test genotypes originated from the traits included in these PCs.

Table 1. Eigenvectors, eigenvalues and percentage of total variance explained by the first five principal components (PC) for 18 traits in 49 tef genotypes.

Traits	Eigenvectors				
	PC1	PC2	PC3	PC4	PC5
Days to seedling emergence(DTE)	-0.11	-0.23	0.28	0.27	-0.50
Days to heading(DTH)	0.33	0.17	-0.05	-0.17	0.15
Days to physiological maturity(DTM)	0.32	-0.23	-0.05	0.00	0.10
Grain filling period (GFP)(days)	0.05	-0.52	-0.01	0.20	-0.04
Plant height (PH)(cm)	0.36	-0.05	0.05	0.08	0.03
Panicle length (PL)(cm)	0.34	-0.05	-0.09	0.11	0.07
Culm length (CL)(cm)	0.32	-0.04	0.16	0.05	-0.01
Peduncle length (PDL)(cm)	0.01	-0.36	0.21	0.09	-0.08
No. of spikelets/panicle(NSPP)	0.29	-0.01	-0.13	-0.11	-0.31
No. of primary main shoot panicle branches (NPMSPB)	0.28	0.17	-0.13	-0.23	-0.24
No. florets/spikelet(NFPS)	0.07	0.25	-0.08	-0.04	-0.72
No. total tillers/plant(NTTPP)	-0.10	-0.40	-0.34	-0.38	-0.06

3.3. Principal Component Analysis for Yield Contributing Traits

Among the 18 traits studied, 10 of them had high contribution effect to the first PC, and these traits included days to heading, days to physiological maturity, plant height, panicle length, culm length, number of spikelets per panicle, number of primary panicle branches per main shoot, lodging index, above-ground biomass and harvest index. The second component predominantly illustrates variation in grain filling period, number of total tillers per plant, number of fertile tillers per plant, days to mature, peduncle length, number of florets per spikelet and thousand-seed weight. The third principal component was chiefly accounted by variation in days to seedling emergence, culm length, peduncle length, lodging index, above-ground biomass yield, grain yield, harvest index, number of total and fertile tillers per plant. The fourth principal component indicated with high variation in grain filling period, fertile tiller per plant, total tiller per plant, days to seedling emergence, lodging index, grain yield, harvest index and thousand-seed weight. The fifth principal component that accounted for about 6 percent of the total variation was due mainly to high variation in days to seedling emergence, number of spikelets per panicle, number of primary panicle branches per main shoot and number of florets per spikelet (Table 1).

3.4. Correlation Among Yield Contributing Traits

The applied method of PCA made it possible to fully assess the relations among tef traits which are used for the analysis of observed diversity in regard to different traits. The most prominent relations shown in Figure 2 are: a strong positive association between GY and SBM; between PL and PH; between GY and PL; among NSPP, CL, PH and PL; among NFTP and NTP; as indicated by the small obtuse angles between their vectors ($r = \cos 0 = +1$). There was a near zero correlation between HI and NFPS, between LI and NFPS, between HI and DTE (Figure 2) as indicated by the near perpendicular vectors ($r = \cos 90 = 0$). There was a negative correlation between LI and GY, and between DFTPPE and NFPS, and NTP and NFPS (Figure 2) as indicated by the angle of approximately 180 degrees ($r = \cos 180 = -1$).

Traits	Eigenvectors				
	PC1	PC2	PC3	PC4	PC5
No. fertile tillers/plant(NFTPP)	-0.13	-0.39	-0.34	-0.36	-0.09
Lodging index (LI) (%)	-0.25	0.07	-0.22	0.27	0.05
Biomass yield(SBM) (kg/ha)	0.28	-0.04	-0.32	0.20	0.07
Grain yield (GY)(kg/ha)	0.06	0.07	-0.56	0.39	0.01
Harvest index (HI) (%)	-0.25	0.13	-0.32	0.26	-0.09
Thousand-seed weight(TSW) (g)	0.16	-0.19	0.04	0.40	-0.05
Eigen values	7.19	2.51	2.28	1.25	1.03
% of variance explained	40.00	14.00	13.00	7.00	6.00
Cumulative% of variance explained	40.00	54.00	67.00	74.00	79.00

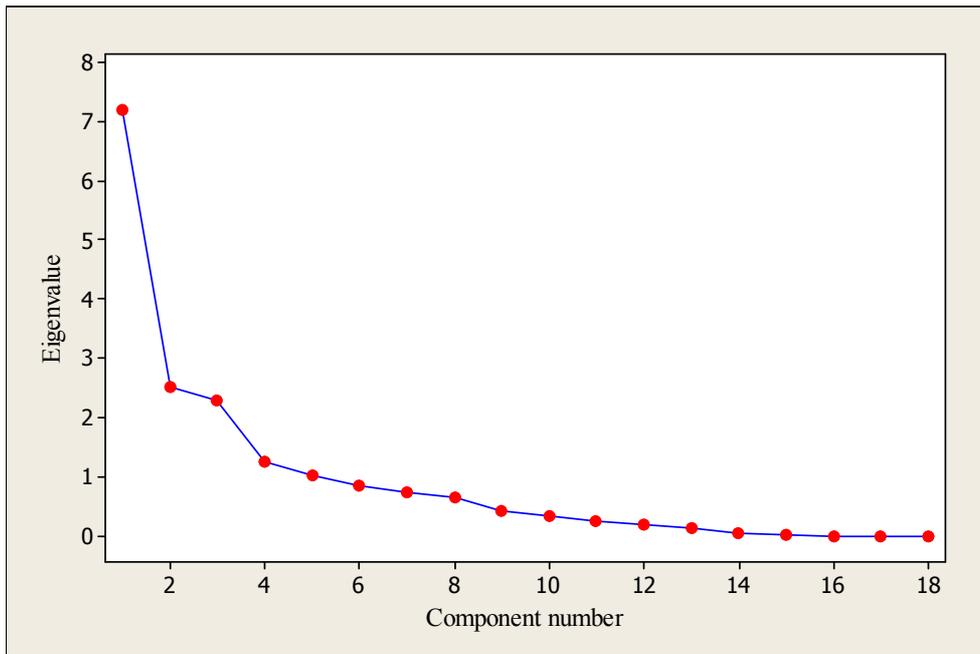


Figure 1. Scree plot of eigenvalue verses component number of 18 quantitative traits in 49 drought tolerant tef genotypes.

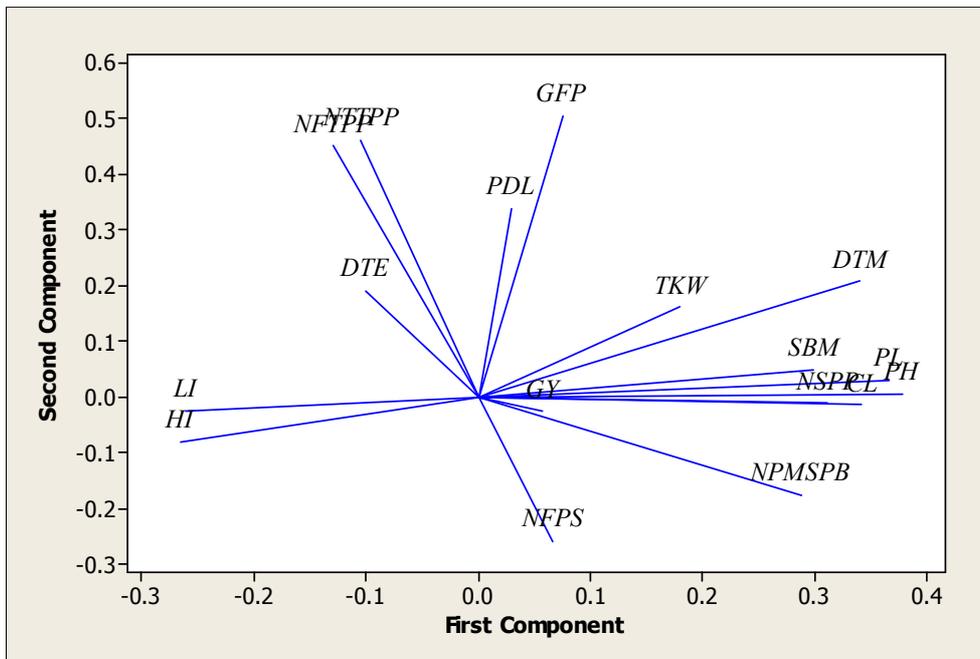


Figure 2. Plot of the first two PCAs showing relation among various tef traits. For explanation of character symbols, see Table 1.

4. Conclusion

We concluded that significant diversity existed among drought tolerant tef genotypes for the traits studied. Five principal components with Eigen-values greater than unity accounted for 79 percent of the gross variability observed for 18 traits across 49 tef genotypes. The first principal component alone explained 40 percent of the total variation, while PC2, PC3, PC4 and PC5 in that order accounted for 14, 13, 7, and 6 percent of the gross observed variation among the test drought tolerance tef genotypes. The PCAs and factor analysis are statistical techniques that are useful for the description of the relations that occur among drought tolerant tef genotypes characteristics. The obtained non-correlated traits may be used for further analysis, where the assumption of having no co-linearity problem of variables is needed.

Acknowledgements

The authors are thankful for the financial help provided by the Ethiopian Institute of Agriculture (EIAR), as well as the technical and material assistance provided by the staffs of DZARC (Debre Zeit Agricultural Research Center) and MARC (Melkassa Agricultural Research Center).

References

- [1] Abdi H., and Williams L. J. 2010. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2, 433-459.
- [2] 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): 00045.
- [3] Alemayehu Muluneh. 2015. Strategies to adapt to climate change in the Central Rift Valley of Ethiopia: landscape impact. PhD Thesis, Wageningen University, Netherland.
- [4] Blum, A. 2011. Plant Breeding for Water-Limited Environments. Springer, New York, USA. pp. 153-216.
- [5] Cannarozzi, G., Solomon Chanyalew, Kebebew Assefa, Abate Bekele, Annett Weichert R. B., Klausner D., Plaza-Wu thrich S., Esfeld K., Jost M., Rindisbacher A., Habte Jifar, Johnson-Chadwick V., Ermias Abate, Wang W., Kamies R., Negussu Husein, Worku Kebede, Kidist Tolosa, Yazachew Genet, Kidu Gebremeskel, Brikti Ferede, Firew Mekbib, Federico Martinelli F., Hans Christian Pedersen H. C., Rafudeen S., Shimelis Hussein, Muluneh Tamiru, Nakayama N., Mike Robinson M., Ian Barker I., Samuel Zeeman S. and Zerihun Tadele. 2018. Technology generation to dissemination: lessons learned from the tef improvement project. *Euphytica*, 214: 31.
- [6] CSA. 2020. Central Statistical Agency, Agricultural Sample Survey 2019/2020 (2012E. C). Volume I. Report on Area and Production of Major Crops (Private Peasant Holdings, Meher Season). Bulletin No. 587 Addis Ababa, Ethiopia.
- [7] Geremew Bultosa, Hall A. N. and Taylor J. R. N. 2002. Physico-chemical characterization of grain tef [*Eragrostis tef* (Zucc.) Trotter] starch, *Starch-Starke*, 54: 461-468.
- [8] Habtamu Ayalew, Tsige Genet, Tadesse Dessalegn, and Landuber Wondale. 2011. Multivariate diversity, heritability and genetic advance in tef landraces in Ethiopia. *African Crop Science Journal*, 19: 201-212.
- [9] 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 and Food Security*, 4:7.
- [10] Holland S. M. 2008. Principal components analysis (PCA). Department of Geology, University of Georgia, Athens, GA, 30602-250.
- [11] Jeffrey, J. 2015. Will Ethiopia's teff be the next 'super grain'? BBC, London.
- [12] Jolliffe, I. T. 2002, Principal Component Analysis, Second Edition *springer verlag* New York, Inc. ISBN 0-387-95442-2.
- [13] Kebebew Assefa, J. K. Yu, M. Zeid, Getachew Belay, HailuTefera and Sorrells M. E. 2011. Breeding tef [*Eragrostistef* (Zucc.) trotter]: conventional and molecular approaches. *Plant Breeding*, 130: 1-9.
- [14] Kebebew Assefa, Seyfu Ketema, Hailu Tefera, Nguyen H. T., Blum A. and Mulu Ayele. 1999. Diversity among germplasm lines of the Ethiopian cereal tef [*Eragrostis tef* (Zucc.) Trotter]. *Euphytica*, 106: 87-97.73.
- [15] Kebebew Assefa, Seyfu Ketema, Hailu Tefera, Tiruneh Kefyalew and Fufa Hundera. 2000. Trait diversity, heritability and genetic advance in selected germplasm lines of tef [*Eragrostis tef*(Zucc.)Trotter]. *Hereditas*, 133: 29-37.
- [16] Mary, S. S., and A. Gopalan. 2006. Dissection of Genetic Attributes Yield Traits of Fodder Cowpea in F3 and F4. *Journal Applications Sciences Researcher* 2 (6): 805–08.
- [17] Mulu Ayele, Blum A. and Nguyen H. T. 2001. Diversity for osmotic adjustment and root depth in tef [*Eragrostis tef* (Zucc.) Trotter]. *Euphytica*, 121: 237-249.
- [18] Mulu Ayele. 1993. Use of excised-leaf water content in breeding tef [*Eragrostis tef* (Zucc.)Trotter] for moisture stress areas. *Acta Agronomica Hungarica*, 42: 261-265.
- [19] Origin Lab. Corporation. 2016. Interpreting Results of Principal Component Analysis.
- [20] Piccinin, D. 2002. More about Ethiopian food: Tef. University of Washington, USA. <http://ethnomed.org/clinical/nutrition/more-about-ethiopian-food-teff>.
- [21] Plaza, S. W., Cannarozzi G. and Zerihun Tadele. 2013. Genetic and phenotypic diversity in selected genotypes of tef [*Eragrostis tef* (Zucc.) Trotter]. *African Journal of Agricultural Research*, 8: 1041-1049.
- [22] Provost, C. and Jobson E. 2014. Move over quinoa, Ethiopian's teff poised to be next big super grain. <http://www.theguardian.com/global-development/2014/jan/23/quinoa-ethiopia-teff-super-grain>.
- [23] SeyfuKetema. 1997. Tef: [*Eragrostistef* (Zucc.) Trotter]: Promoting the Conservation and Use of Underutilized and Neglected Crops12 Institute of Plant Genetics and Crop Plant Research, Gatersleben/International Plant Genetic Resources Institute, Rome, Italy.
- [24] Spaenij-Dekking, L., Kooy-Winkelaar Y. and Koning F. 2005. The Ethiopian cereal tef in celiac disease. *The New England Journal of Medicine*, 353: 1748-1749.

- [25] Tsion Fikire. 2016. Variability Genetic Diversity of Ethiopian Tef (*Eragrostis tef* (Zucc.) Trotter) Varieties as Revealed by Morphological and Microsatellite Markers. MSc. Thesis, Addis Ababa University, Addis Ababa, Ethiopia.
- [26] Vavilov, N. I. 1951. The origin, variation, Immunity and Breeding of cultivated plants. Ronald Press, New York. [Translated from Russian by Sarrchester, K].
- [27] Yan, W., and M. S. Kang. 2003. GGE Biplot Analysis: A Graphical Tool for Breeders, Geneticists, and Agronomists. Boca Raton, FL: CRC Press.
- [28] Zerihun Tadele and Kebebew Assefa. 2012. Increasing food production in Africa by boosting the productivity of understudied crops. *Agronomy*, 2: 240–28.