ScancePG
Science Publishing Group

Genetic Diversity Based on Cluster and Principal Component Analyses for Agro-morphological Traits of Wheat Germplasm

Solomon Mengistu[†], Mekonnen Asefa[†]

Crop and Horticulture Directorate, Ethiopian Biodiversity Institute, Addis Ababa, Ethiopia

Email address:
mengistusolomon7@gmail.com (Solomon Mengistu), hunem2018@gmail.com (Mekonnen Asefa)
† Solomon Mengistu and Mekonnen Asefa are co-first authors.

Abstract: Understanding of the genetic variability in germplasm is crucial for successful choosing varsities in breeding programs. An investigation was conducted to evaluate the genetic variability by cluster and principal component analysis for yield and its nine contributing traits in twenty seven wheat accessions that obtained from Ethiopian Biodiversity Institute and two standard checks namely: - Hidase and Denda were planted at Melkasa Agricultural Research Farm 2017, in main rainy season. Out of different techniques available for assessing the genetic diversity, principal component and cluster analysis are the most important and widely used methods The agro-morphological characters used for analysis were days to 50% heading, plant height, number of tillers per plant, days to 75% maturity, spike length, number of spikelets per spike, number of grain per spike, grain size and yield per plot. Principal component analysis categorized the entire accessions into nine important components explaining nearly 100% of genetic variation. Cluster analysis similarly, categorized the examined characters into four groups. In cluster IV, two accessions namely AC-231520 and AC-222300 have achieved higher rates of yield and yield attributes traits and may be reflected as potential to advance breeding programs. The outcome of the current research could be utilized in designing and implementation of forthcoming breeding programs in wheat.

Keywords: Characters, Germplasm, Principal Component and Cluster Analysis

1. Introduction

In Ethiopia wheat (*Triticumspp*) is supposed a tactical food security crop in terms of land assigned, amount produced, households evolved and income generation. The annual production of the country is estimated 5.8 million tons with average productivity of 3 tons per hectare [1], which is comparatively smaller than the achievable yield of the crop, getting up to 5 tonsper hectare [2]. Wheat shares about 17% of entire grain production in Ethiopia making it the third essential cereal crop followingteff (*Eragrostistef*, Zucc) Trotter] and maize (*Zea mays L.*). The suitable elevation for wheat production is found in the range of 1900 to 2700 meters above sea level [3].

In Ethiopia, wheat production has been affected by complicated biotic and abiotic causes including erratic rainfall, poor management practices and depletion of soil fertility, diseases and insect pests. Increasing wheat production without extensive agriculture implies that we must increase wheat production on existing agricultural lands through application of anintensive farming system. This could be attained partially by improving wheat yield genetic potential through a better investigation of characters involved and their relationship with the environment. Tahmasebi *et al.*, [4] investigated wheat breeding targeting onthe improvement of wheat cultivars by using superior traits to boost the yields of the crop in various environments. The achievementof a successful breeding program relies on the investigation of genetic differences among traits and relationships between unlike characters that ultimately provides theyield of a crop [5]. The isolation of superior genotypes is mainly depending on the potential of

cultivarsand able to increase yield under different environmental conditions. Genetic differences and heritability of cultivars are crucial elementsin influencing yield and yield attributes traits [6]. Different techniques are existed for investigation of crop genetic variability and a number of researchers argued that cluster and principal component analyses are important to examine genetic variability and interaction of wheat genotypes [7-9].

Brown-Guedira, G. L. *et al*, [10] cluster analysis (CA) is important in grouping samples possessed multipart characters rather than single character. Principal component analysis has the capacity to change more number of associated parameters into a lesser number of class variables called principal components. Mujaju *et al*. [11] reported principal component analysis (PCA) could be carried on prior to cluster analysis (CA). Principal component analysis (PCA) gives opportunity for the identification of promising sources for conversional breeding purposes [12]. The present study was undertaken to assess the extent of genetic variability and traitsprovidingtowards total genetic variability in some wheat germplasm for its exploitation in future wheat breeding programs.

2. Material and Methods

The current study was conducted at the experimental field of Kulumsa Agricultural Research Center (KARC), Oromia, Ethiopia. It is located at latitudeof 8°2'N and longitude of 39°10'E in central Ethiopia. It has an average elevation of 2200m above sea level. The region has a dark reddish brown with PH values of 6.6. The annual rainfall of the location is measured at 840mm with 10°C and 22°C mean daily minimum and maximum temperature. Twenty seven wheat germplasm and two standard checks namely: Hedasse and Dandaha which were obtained from Kulumsa Agricultural Research Center (KARC) were used for yield and yield component characterscomparison. Twenty seven wheat germplasm without replication along with two replicated checks were arranged in an Augmented Design. Individual plot size was measured 2m*0.80m with 4 rows and planted at a spacing of 23*10cm. The other recommended agronomic management practices were applieduniformly. Twenty randomly selected plants from each plot were used to record data on days to 50% heading, plant height (cm), number of tillers per plant, days to 75% maturity, spike length, number of spikelets per spike, number of grains per spike, grain size (mm) and yield per plot (g). The data collected for all quantitative characters were subjected to the analysis of basic statistics; correlation, clusterand principal component analysis using the Statistical Package for the Social Sciences (SPSS) 16.0 software [13].

3. Results and Discussion

The results of calculated statistics for the investigated traits is displayed in Table 1 that maximum genetic variability was recorded for spike length (28.48%), followed by thenumber of grain per spike (24.29%) and yield per plot (20.15%). Conversely, the lower values of genetic differences were accounted for days to heading (9.69%), kernel size (11%) and plant height (11.28%). The results of the statistical analysis confirmed that there was much variability among the tested germplasm (Table 1). Similarly, (Drezner *et al.,* [14] argued that yield of grain is the result of various characteristics which is the outcome of yield attributes characters like the amount of spikelets per spike, weight of thousand seeds, height of the plant and additional characters also affected by varieties and climate situation.

Table 1. Basic statistics for various characters of wheat germplasm.

Characters	Mean	Minimum,	Maximum,	Range,	SD	CV (%)
Days to heading	66.10	58.00	86.00	28.00	6.41	9.69
Plant height (cm)	97.27	71.50	124.00	52.50	10.97	11.28
Number of tillers per plant	1.68	1.10	2.60	1.50	0.31	18.66
Days to maturity	120.31	88.00	144.00	56.00	17.03	14.15
Spike length (cm)	9.14	6.40	21.30	14.90	2.60	28.48
Number of spikelet per spike	18.45	11.40	33.40	22.00	3.65	19.77
Number of grain per spike	35.53	4.40	45.80	44.1	8.63	24.29
Grain size (mm)	6.46	4.60	7.60	3.00	0.71	11.00
Yield per plot (g)	436.52	291.00	597.00	306.00	87.95	20.15

SD = Standard deviation, CV = Coefficient of variation.

3.1. Correlation Analysis

Phenotypic and genotypic association between and among differenttraits are displayed in Table 2. Grain yield hadaposutive relationship with days to 50% heading (0.34), height of plant (0.09), number of tillers per plant (0.18), days to 75% maturity (0.18) and spike length (0.04). However it had negative associations among number of spikelets per spike, number of grains per spike and grain size (Table 2). Similarly, positive correlation of grain yield and plant height was reported byLeilah, A. A. and Al-Khateeb, SA. [15]. In addition, different researchers confirmed that negative relationship between yield and yield attributes traits was observed in wheat germplasm characterization [16].

Table 2. Correlation analysis for agro-morphological traits evaluated for wheat germplasm traits.

Character	DH	PH	NT	DM	SL	NSPS	NGPS	GS	YPP
DH		0.18	0.14	0.84**	-0.25	-0.30	-0.06	-0.22	0.34
PH			0.00	0.42*	0.03	0.20	-0.15	0.03	0.09
NTPP				-0.02	-0.07	-0.17	0.03	0.19	0.18
DM					-0.22	-0.15	-0.03	-0.22	0.18
SL						0.79**	-0.55*	0.24	0.04
NSPS							-0.31	0.45*	-0.12
NGPS								0.18	-0.25
GS									-0.26
YPP									

*Significant at P = 0.05, ** Significant at P = 0.01, DH= Days to heading, PH= Plant height, NTPP=Number of tillers per plant, DM= Days to maturity, SL= Spike length, NSPS= Number of spikelet/spike, NGPS= Number of grain per spike, GS= Grainsize and YPP= Yield per plot.

3.2. Principal Component Analysis (PCA)

Principal components analysis is typicallyimportant to make a large whole data into simpler contents. The mean of the data were analyzedwith principal component analysis. This type of analysis shows the total differenceby the side of alignment of isolation. The results of eigenvalues are mainly important to identify the eigenvectors that reflects population parents.

The total of the eigenvalues is normallyequivalent to the amount of variables [17]. Based ont he findings of Chekalin, *et al* [18] traits possessing positive rates and nearer to joining together occurred in the first principal component and affects grouping genotypes into clusters. Althoughthe tested 29 wheat germplasm yields and yield attributes traits data were calculated by principal component analysis and grouping them in to nine components, four had eigenvalues greater than one. These nine principal components explained about 100% of the entire variability (Table 3). This is reliable to the work by Degewione *et al.,* [19].

Based on the results of data analysis, potential traits for related components were recognized for every of the major components. The statistical analysis is indicating the variation in the population characters. Table 3 shows that PC-1 explained 29.06%, PC2, PC3, PC4, PC5, PC6, PC7, PC8 and PC9 accounted for 22.54%, 14.63%, 13.15%, 8.07%, 6.76%, 3.60%, 1.37% and 0.84% respectively of the total variance among different yield and yield attributes traits.

First principal component had a positive loading with days to 50% heading, plant height, days to 75% maturity, number of tillers per plant and number of grains per spike, while negative loading for spike length, number of spikelets per spikeand grain size. The first principal component described phonological traits (Table 4).

Second principal component had a positive loading with days to 50% heading, plant height, days to 75% maturity, spike length, number of spikelets per spike and yield per plot. Likewise, negative effects were recordedfor number of tillers per plant, number of grains per spike and grain size.

Third principal component had a positive loading with plant height, days to 50% heading, number of tillers per plant, days to 75% maturity, number of spikelets per spike, number of grain per spike, grain size and negative loading with spike length and yield per plot.

Fourthprincipal component had positive loading with plant height, number of tillers per plant, spike length, grain size and yield per plot, while hadnegative loading with plant height, days to 75% maturity, number of spikelets per spike and number of grains per spike.

Fifth principal component had negative loading with plant height, number of tillers per plant and yield per plot, while positive loading with days to 50% heading, days to 75% maturity, spike length, number of spikelets per spike, number of grain per spike and grain size.

Sixth principal component had negative loading ondays to 50%heading, plant height, number of tillers per plant, days to 75%maturity and yield per plot and positive loading with spike length, number of spikelets per spike, number of grain per spike and grain size.

Seventh principal component had a positive loading on thenumber of tillers per plant, days to 75% maturity, spike length, number of spikelets per spike and number of grain per spike and had negative loading with days to 50% heading, plant height, grain size and yield per plot.

Eighth principal component had a positive loading with days to 50%heading, plant height, spike length, number of grain per spike and grain size and had negative loading on number of tillers per plant, days to maturity, number of spikelets per spike and yield per plot.
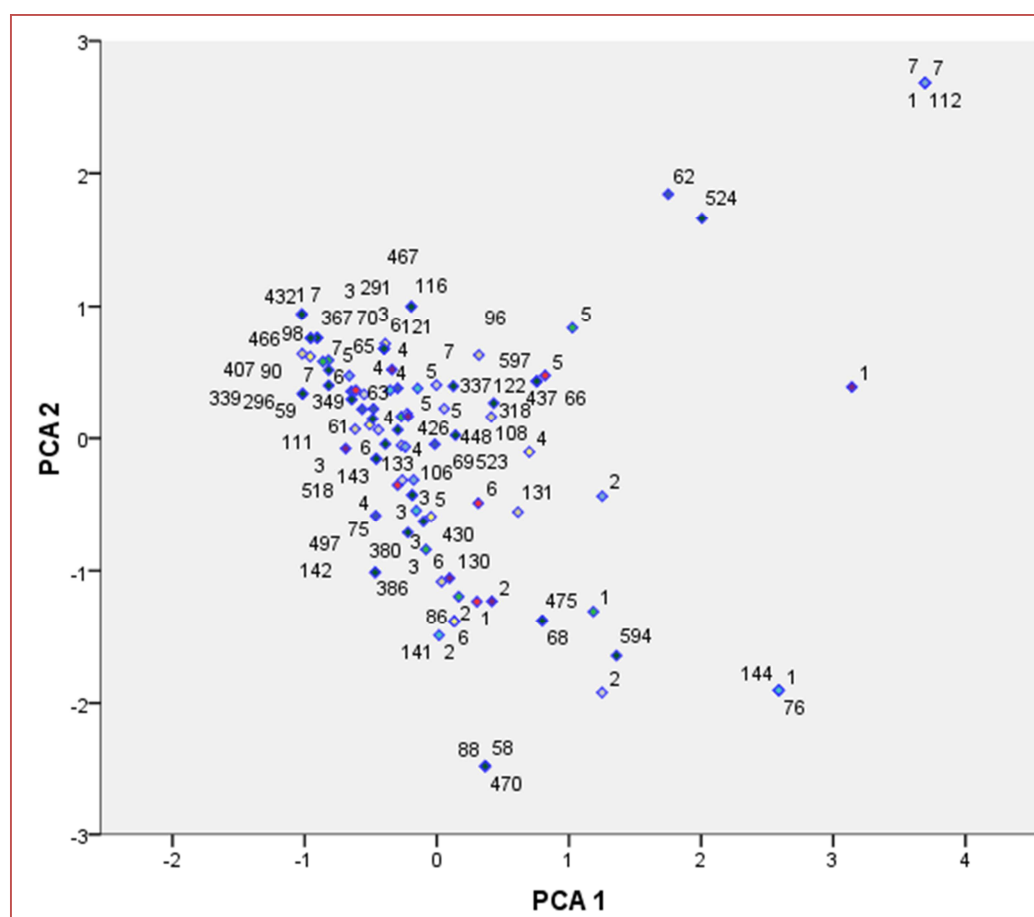
Ninth principal component had a positive loading on days to 50%heading, plant height, number of spikelets per spike and number of grain per spike and had negative loading with number of tillers per plant, days to 75%maturity, grain size and yield per plot (Table 4). Principal component analysis is typicallyappliedfor making variables smaller and arranged genotypes in groups. Therefore principal component analysis is one of tools used to identify ancestral lines for a successful breeding program [20].

Table 3. Principal component analysis of different characters on wheat germplasm.

Character	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eigen value	2.62	2.03	1.32	1.18	0.73	0.61	0.32	0.12	0.08
% of total variance	29.06	22.54	14.63	13.15	8.07	6.76	3.60	1.37	0.84
Cumulative variance%	29.06	51.60	66.23	79.38	87.45	94.20	97.80	99.16	100

Table 4. *Factors loading for yield and yield attributes characters of 29 wheat germplasm.*

Character	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
DH	0.47	0.31	0.16	0.06	0.45	-0.01	-0.08	0.16	0.64
PH	0.08	0.38	0.39	-0.21	-0.76	-0.04	-0.03	0.20	0.16
NTPP	0.08	-0.03	0.24	0.80	-0.13	-0.40	0.32	-0.11	-0.02
DM	0.43	0.36	0.29	-0.21	0.25	-0.06	0.18	-0.15	-0.66
SL	-0.45	0.40	-0.11	0.08	0.22	0.19	0.31	0.68	-0.13
NSPS	-0.47	0.35	0.19	-0.10	0.12	0.48	0.35	-0.62	0.26
NGPS	0.13	-0.50	0.39	-0.09	0.01	0.23	0.54	0.23	0.06
GS	-0.31	-0.05	0.62	0.23	0.19	0.72	-0.59	0.07	-0.15
YPP	0.21	0.30	-0.32	0.43	-0.18	-0.02	-0.06	-0.05	-0.11

Figure 1. *Distribution of wheat accessions for the first two principal components.*

3.3. Cluster Analysis

Cluster analysis isolates genotypes into clusters which have demonstrated closer similarities within a cluster and dissimilarities among clusters. The current investigation was separated the twenty nine wheat accession into four clusters based on various yield and yield attributes traits such as days to 50% heading (DH), plant height (PH), number of tillers/plant (NTPP), days to 75% maturity (DM), spike length (SL), number of spikelets/spike (NSPS), number ofgrains/spike (NGPS), grain size (GS) and yield/plot (YPP).

The results of all clusters were interpreted by considering their average values (Table 6). The average rates for number of grain per spike (38.24) and grain size (6.67) were greatest in cluster I. Cluster II accounted the higher rates for plant height (104.5), spike length (10.92), number of spikelets per spike (20.07). Cluster III be composed of genotypes which have strongly resemblance to cluster I. In addition, cluster IV accounted forthe greatest average rates for days to 50% heading (76.0), number of tillers per plant (1.83), days to maturity (135.5) and yield per plot (595g). In this group, two accessions namely AC-231520 and AC-222300 were found superior in their greater rates of yield and yield attributes traits and might be take into account as potential to further breeding program. Ajmal, *et al*., [21] argued that cluster analysis makes the isolation procedure simple by considering genotypes recorded higher values yield attributes traits and grouped in single cluster.

Table 5. Clustering pattern of the 29 wheat germplasm based on agro-morphological characters.
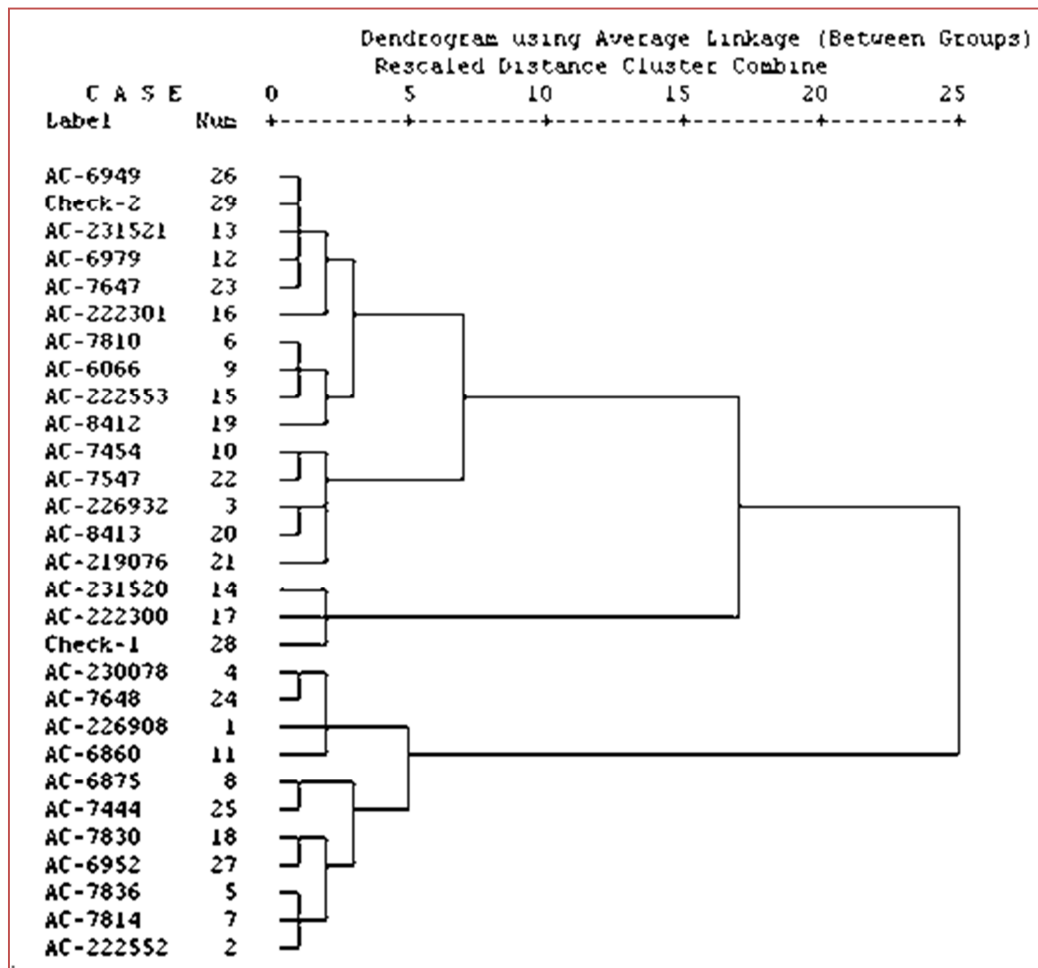
Cluster name	No. of germplasm	Name of accessions in each cluster					
I	11	AC-226908	AC-222552	AC-230078	AC-7836	AC-7814	AC-6875
		AC-6860	AC-7830	AC-7648	AC-7444	AC-6952	
II	5	AC-226932	10:AC-7454	AC-8413	AC-219076	AC-7547	
III	10	AC-7810	AC-6066	AC-6979	AC-231521	AC-222553	AC-222301
		AC-8412	AC-7647	AC-6949	Check 2		
IV	3	AC-231520	AC-222300	Check 1			

Table 6. Cluster means values for agro-morphological characters of 29 wheat germplasm.

Characters	Clustername			
	I	II	III	IV
Days to heading	64.22	65.8	65.09	76.0
Plant height (cm)	97.48	104.5	93.44	96.0
Number of tillers per plant	1.64	1.62	1.71	1.83
Days to maturity	116.7	120.3	119.0	135.6
Spike length (cm)	8.71	10.92	9.09	7.07
Number of spikelet/spike	18.98	20.07	17.98	15.33
Number of grain per spike	38.24	28.45	37.64	33.87
Grain size (mm)	6.67	6.27	6.48	6.13
Yield per plot (g)	334.89	510.17	436.27	595.0

Dendrogram can be used to assess the cohesiveness of the clusters formed and can provide information about the appropriate number of clusters to keep. This procedure attempts to identify relatively homogeneous groups of cases (or variables) based on selected characteristics such as yield and yield attributing (Figure 2). In this study, cluster analysis exhibited that cluster I, II, III, and IVcontained 11, 5, 10 and 3 accessions respectively. Similarly, cluster analysis might be important for the selection of superior yielder wheat varieties as reported by Ali, Y. *et al.*, [22].

Figure 2. Tree diagram of 29 wheat accessions based on different yield characters.

4. Conclusion

The outcome of PC analysis showed the major components that supplied to a great extent to the assessment of higher yielder wheat genotypes. Thus, in the first and second components, days to 75% maturity had a positive and maximum loading and yield per plot had the highest loading for the fourth component. In cluster analysis, twenty nine wheat accessions were grouped into four clusters and the average rates for number of grain per spike (38.24) and grain size (6.67) were recorded the maximum values in cluster I. Cluster II accounted the higher rates for plant height (104.5), spike length (10.92), number of spikelets per spike (20.07). Cluster III be composed of genotypes which have strongly resemblance to cluster I. In addition, cluster IV accounted for the greatest average rates for days to 50% heading (76.0), number of tillers per plant (1.83), days to maturity (135.5) and yield per plot (595g). Therefore, the present study provides vital information for wheat variety development through conventional breeding programs.

References

[1] CSA. 2021. Agricultural sample survey: Report on area and production of major crops (Private peasant holdings, Meher Season). Volume I Statistical Bulletins 590, Addis Ababa, Ethiopia.

[2] Zegeye F, Alamirew B, Tolossa D. 2020. Analysis of wheat yield gap and variability in Ethiopia. Int J AgricEconon. 5 (4): 89-98.

[3] Hundie B, Verkuiji H, Mawangi W, Tanner DG. 2000. Adaptation of improved wheat technologies in adaba and dodolaworedas of the Bale highlands of Ethiopia. Addis Ababa, Ethiopia: CIMMYT/EARO.

[4] Tahmasebi, G., J. Heydarnezhadian and A. P. Aboughadareh. 2013. Evaluation of yield and yield components in some of promising wheat lines. Inter. J. Agric. Crop Sci., 5 (20): 2379-2384.

[5] Mary, S. S. and A. Gopalan. 2006. Dissection of genetic attributes yield traits of fodder cowpea in F3 and F4. *J. Appl. Sci. Res*., 2 (6): 805-808.

[6] Baenziger, P., I. Dweikat, K. Gill, K. Eskridge, T. Berke, M. Shah and A. Mahmood. 2011. Understanding grain yield: It is a journey, not a destination. C*zech. J. Genet. Plant Breed.* 47 (Special Issue): S77-S88.

[7] Devesh, P., Moitra, P. K., Shukla, R. S. and Pandey, S. (2019) Genetic Diversity and Principal Component Analyses for Yield, Components and Quality Traits of Advanced Lines of Wheat. Journal of Pharmacognosy and Phytochemistry, 8, 4834-4839.

[8] Beheshtizadeh, H., Rezaie, A., Rezaie, A. and Ghandi, A. (2013) Principal Component Analysis and Determination of the Selection Criteria in Bread Wheat (*Triticumaestivum* L.) Genotypes. The International Journal of Agriculture and Crop Sciences, 5, 2024-2027.

[9] Lysenko, A. A. (2011) Comparative Productivity of Pea Varieties of Various Morphotypes and the Creation of a New Selection Material on Their Basis. Zernograd, 23.

[10] Brown-Guedira, G. L., Thompson, J. A., Nelson, R. L. and Warburton, M. L. (2000) Evaluation of Genetic Diversity of Soybean Introductions and North American Ancestors Using RAPD and SSR Markers. Crop Science, 40, 815-823. https://doi.org/10.2135/cropsci2000.403815x

[11] Mujaju, C. and Chakuya, E. (2008) Morphological Variation of Sorghum Landrace Accessions On-Farm in Semi-Arid Areas of Zimbabwe. International Journal of Botany, 4, 376-382. https://doi.org/10.3923/ijb.2008.376.382

[12] Mustafa, H. S., Farooq, J., Ejaz-ul-Hasan, Bibi, T and Mahmood, M. (2015) Cluster and Principle Component Analyses of Maize Accessions under Normal and Water Stress Conditions. Journal of Agricultural Sciences, 60, 33-48. https://doi.org/10.2298/JAS1501033M

[13] L evesque, R. (2007). SPSS. Statistical package for the social sciences, version 16.0. SPSS Programming and data management. A guide for SPSS and SAS users, fourth edition, SPSS inc., Chicago III.

[14] Drezner G, Dvojkovic K, Horvat D, Novoselovic D, Lalic A. 2007. Environmental impacts on wheat agronomics quality traits. Cereal Research 35, 357–360.

[15] Leilah, A. A. and Al-Khateeb, SA. (2005). Statistical analysis of wheat yield under drought conditions. Journal of Arid Environments, 61 (3): 483-496.

[16] Deng, S., X. Wu, Y. Wu, R. Zhou, H. Wang, J. Jia and S. Liu. 2011. Characterization and precise mapping of a QTL increasing spike number with pleiotropic effects in wheat. *Theor. Appl. Genet.,* 122 (2): 281-289.

[17] Hailegiorgis, D., Mesfin, M, and Genet, T. (2011) Genetic Divergence Analysis on Some Bread Wheat Genotypes Grown in Ethiopia. Journal of Central European Agriculture, 12, 344-352. https://doi.org/10.5513/JCEA01/12.2.922

[18] Chekalin, N. M., Tishchenko, V. N. and Batashova, M. E. (2008) Selection and Genetics of Individual Cultures. FOP Govorov, S. V., Poltava, 368 p.

[19] Degewione, A. and Alamerew, S. (2013) Genetic Diversity in Bread Wheat (*Triticumaestivum*L.) Genotypes. Pakistan Journal of Biological Sciences, 16, 1330-1335. https://doi.org/10.3923/pjbs.2013.1330.1335

[20] Akter, A., Hasan, M. J., Paul, A. K., Mutlib, M. M. and Hossain, M. K. (2009) Selection of Parent for Improvement of Restorer Line in Rice (*Oryza sativa* L.). SAARC Journal of Agriculture, 7, 43-50.

[21] Ajmal, S. U., Minhas, N. M., Hamdani, A., Shakir, A., Zubair, M., Ahmad, Z. 2013. Multivariate analysis of genetic divergence in wheat (*Triticumaestivum L*) germplasm. Pak. J. Bot., 45 (5): 1643-1648.

[22] Ali, Y., Atta, B. M., Akhter, J., Monneveux, P., Lateef, Z. 2008. Genetic variability, association and diversity studies in wheat (*Triticumaestivum L.)* germplasm. Pak. J. Bot., 40 (5): 2087-2097.