Journal of Genetics, Genomics & Plant Breeding 5(4) 86-94 (October, 2021)
ISSN (Online): 2581-3293

RESEARCH ARTICLE

Multivariate analysis of cotton genotypes based on agronomic traits

Martin Lubinda Simasiku[1], Ejiet John Wasige[2], Langa Tembo[3]

1 Cotton Development Trust, Department of Cotton Breeding, P.O Box, 670057, Mazabuka, Zambia

2 Makerere University, Department of Agriculture production, P.O Box, 7062, Kampala Uganda

3 The University of Zambia, Department of Plant Science P.O Box 32379, Lusaka, Zambia

Corresponding authors email Id: martinsimasiku@gmail.com

Manuscript received: August 17, 2021; Decision on manuscript, August 30, 2021; Manuscript accepted: October 1, 2021

Abstract

Cotton is an important cash crop in the world and is mainly grown for its fibre in Zambia. The objective of this study was therefore to cluster the cotton genotypes into distinctive grouping. Thirty genotypes were planted in an incomplete block design replicated three times, in seven sites. Several agronomic traits were recorded and mean performance noted. Data analysis using principle components revealed that the parameters, number of bolls and seed cotton yield (SCY) with loading scores of 0.52 and 0.51 respectively, were the best at discriminating genotypic performance. Implying that where resources are limiting, genotypic selection can only be employed based on number of bolls and SCY. In this study, the most dissimilar paired parental genotypes were identified as MG27 (from cluster group A) and MG5 (from cluster group C) with a similarity value of 29.7 %. This parental cross (MG27 X MG5) is expected to create maximum genetic variability among offspring's, creating a wider spread of choice in selecting for desirable genotypes for release or being used as parents in other crosses.

Key words: Cluster analysis, cotton, fiber, seed yield, variability

Introduction

Cotton (*Gossypium hirsutum*) is grown in many parts of the world especially the tropics and temperate regions mainly for its fibre (Egbuta et al. 2017). India is the largest cotton producer in the world with an annual production of approximately 6.4 million tonnes. Cotton is a source of seedcake for animal feed, oil and also a reliable source of income (FAO, 2018). In Zambia, the cotton varieties are preferred for their heat and disease tolerance, but have very low seed cotton yield (SCY) of between 300 to 600 kg/ha as compared to the potential of 2000kg to 3500 kg/ha (Cotton Development Trust, 2015). Its production is affected by many factors both biotic, mostly pests and abiotic conditions such as rainfall, soil fertility and to some extent genetic degeneration.

Critical to genetic improvement is the creation or identification of genetic variability among the germplasm (Dhivya *et al.,* 2013). Characterization of cotton germplasm is a vital tool in the selection of potential parents for development of subsequent desirable hybrids and selection of superior progenies arising from genotypic cross advancement (Murtaza *et al.*, 2005). In this regard, the delineation of germplasm into different genetic groupings allows for genotypes which perform similarly to be grouped together into clusters, to allow the choice of potential parents. Molecular marker and phenotypic trait analysis, known as multi-variate analysis have been employed in characterization and clustering genotypes into distinct groups (Tembo and Munyinda, 2015; Asha *et al.*, 2013). Use of molecular marker analysis are preferable being that they are independent of the environment and hence considered as a more efficient approach (Mbwando *et al.*, 2016). However, where molecular markers are unavailable or inaccessible, use of multi-variate trait analysis for phenotypic characterization and delineation of germplasm is an option. It should be noted that the accuracy of obtaining reliable cluster grouping depends on the efficiency of phenotypic scoring including for cotton as an agronomic trait response is influenced by the environmental effect (Khan *et al.,* 2018). Thus, the use of mean score trait values across environments offers more reliable mean data for multivariate analysis especially for quantitative traits.

Apart from clustering genotypes, multi-variate analysis has been used in identifying traits that best discriminate the genotypes within the same species. Knowing such traits is important in breeding as it helps the breeder to minimise costs where funds are limiting by choosing fewer and appropriate traits as an aid to genotypic selection. In Zambia, the Cotton Development Trust (CDT) has developed and released a number of varieties and also in possession of introduced genotypes (Simasiku *et al.*, 2020). Though their performance across several environments has been established (Simasiku et al., 2020), their genetic similarity and cluster grouping is still unknown. Therefore, the objectives of this study were to identify traits which best discriminate the cotton genotypes and to cluster cotton genotypes into distinctive grouping.

Materials and methods

Experimental layout and site

This study was undertaken in seven sites of Zambia namely: Magoye, Masumba, Liempe, Mutanda, Minsanfu, Msekera and Gwembe (Table 1). Thirty (30) genotypes (Table 2) were planted in an incomplete block design and replicated three times in all the seven sites as by Simasiku *et al.*, 2020. Plants were established in two-row plots at a spacing of 90cm by 30cm in 4-metre long rows. All recommended management and agronomic practices were followed.

Table 1: Experimental sites used during the cropping season

Location	**Coordinates**	**Altitude (m)**	**Soil Type**
Liempe	15°22’S, 28°26’E	1171	Sandy loam
Magoye	15°59’S, 27°37’E	1018	Sandy clay loam
Gwembe	16°29’S, 27°35’E	534	Sandy Clay
Msekera	13°38’, 32°34’ E	1032	Sandy loam
Masumba	13°22’S, 31° 56’E	546	Loamy sand
Mutanda	12°25’S, 26°12’E	1300	Sandy loam
Misamfu	10°17’S, 31°22’ E	1536	Sandy clay loam

Table 2: Germplasm used in the multivariate cluster analysis during the cropping season

Genotype Code	Genotypic Pedigree	Genotype Code	Characteristic Growth Habit
M G1	BC4 x CDT II	C1104	Indeterminate
MG2	BC4 x CDT V	C1105	indeterminate
MG3	CDT-09 x BP 52	C1112	indeterminate
MG4	CDT II x Turk A	C1109	indeterminate
MG5	Rocket x CA336	C2612	indeterminate
MG6	Cameroon A x Zim II	C1107	indeterminate
MG7	MF20kG x VH8 4620	C2614	indeterminate
MG8	BC1 x C2511	C1103	indeterminate
MG9	CA347 x F135	C2602	indeterminate
MG10	C457 x CA336	C2619	indeterminate
MG11	Rocket x G319-18	C2618	indeterminate
MG12	CDT II x Turk B	C1110	indeterminate
MG13	CDT II x BP 52	C1111	indeterminate
MG14	BC4 x ISC 4	C1101	indeterminate
MG15	Ihmad 742 x Chureza	C1116	indeterminate
MG16	CA223 x CDT V	C1114	indeterminate
MG17	CA223 x CDT II-09	C1113	indeterminate
MG18	Stam29ABG1818 x CDT II-09	C1106	indeterminate
MG19	Cameroon A x Zim III	C1108	indeterminate
MG20	Turk B x BP52	C1119	indeterminate
MG21	CDT II-06 x Cameroun A	C1115	indeterminate
MG22	Turk B x Cameroun A	C1120	indeterminate
MG23	BC 3 x ISC 6	C1102	indeterminate
MG24	Cameroun A x BP 52	C1121	Determinate
MG25*	MV 513 x MV515	C 567	Determinate
MG26*	MV513 x MV 517	C571	Determinate
MG27*	MV513 xMV516	C 569	Determinate
MG28	(G319-16xcza87)x(BIII-F3xG319-16)	Cotton Development Trust II	Determinate
G29	CA336	Cotton Development Trust V	Determinate
MG30	C1188 x L299)	Chureza	indeterminate

Genotypes MG 25, MG26 and MG 27 are F_1 hybrids obtained from Mahyco, while the rest are lines obtained from Cotton Development Trust

Data collection and analysis

Data for all measured parameters was collected for all the seven sites and mean values for each parameter across sites was recorded. Collection of the SCY was done at 50 % boll opening and at harvest using a digital weighing scale. Counting of the number of open bolls was carried out at harvest. Plant height was measured when the plants were approximately 130 days after germinating, using a tape meter. The ginning out-turn (GOT) was evaluated after harvest using a laboratory ginning machine. The GOT was computed as a ratio of the amount of lint over the total seed cotton weight and expressed as a percentage. Seed index, which is the weight of 100 cotton seeds per genotype was also measured. Multi-variate analysis, utilizing the means of all measured parameters were undertaken using principal component analysis (PCA) to determine the most discriminating parameter and to schematically apportion the genotypes in distinct groups. The dendrogram and similarity matrix was finally constructed using cluster analysis. All data analysis was performed using GenStat statistical software 18th edition (Payne *et al.*, 2010).

Results and discussion

Mean parameter measurements across locations

The mean performance of SCY, GOT, Plant height, seed index and number of bolls were recorded (Table 3). The standard error of the mean was computed as 45.6 kg/ ha, 0.3 %, 2.1 cm, 0.1 g and 9.4 bolls respectively. The standard error values indicate that several genotypic mean performances for all measured parameters fell above or below the grand genotypic (population) mean.

Multivariate evaluation of genotypes

Evaluation of principle components and measured parameters

Analysis on eigen values (Figure 1) showed that two principle components (PC),1 and 2 contributed most to variations in genotypic responses. This was represented as root 1 and 2 with eigen values of 3.5 and 1 respectively. PC1 and PC2 contributed 68.5 % and 21.5 % respectively giving a total of 90 % of percentage variation explained (Table 4).

The parameters, number of bolls and SCY exhibited higher contribution (greater than 0.5) in differentiating genotypic responses with a loading score of 0.52 and 0.51 respectively. Arising from the fact that PC 1 had a much higher contribution of the percentage variation explained (68.5 %). This was followed up by GOT and seed index arising from PC2 with a loading score of 0.77 and 0.53 respectively. The other three Principle components (PC 3 to 5) had very low percentage variation (approximately 10 %) to be considered as a reasonable contribution.

The discrimination of parameters (traits) is an important aspect to a breeder because it makes him/ her aware of what traits are most important in screening of the candidate genotypes (Evgenidis et al., 2011). Generally screening of germplasm may be costly and identification of important traits may help the breeder to narrow down to a few parameters or traits as an aid to selection.

Table 3. Mean performance of measured parameters across locations

Genotype	Seed cotton yield (Kg/ha)	Ginning out turn (%)	Height (cm)	Seed index (g)	Number of bolls
MG1	345.21	42.38	62.63	10.76	79.19
MG10	195.30	40.40	41.79	10.24	50.76
MG11	208.60	40.66	41.86	10.48	29.38
MG12	386.60	41.59	72.54	10.75	80.25
MG13	404.43	41.93	51.77	9.76	72.10
MG14	396.04	43.29	65.46	10.67	96.24
MG15	313.69	42.94	68.70	10.38	80.48
MG16	440.87	43.02	65.54	10.86	102.71
MG17	479.56	43.06	67.77	10.29	83.62
MG18	724.07	44.30	70.42	11.00	128.00
MG19	648.88	42.90	73.50	10.38	129.19
MG2	439.88	42.20	62.26	10.90	94.14
MG20	546.23	42.88	64.31	10.10	114.33
MG21	568.72	42.71	70.83	10.29	118.48
MG22	497.16	42.43	66.68	10.57	127.38
MG23	678.17	42.90	69.22	11.14	112.38
MG24	516.14	41.26	69.44	10.86	136.95
MG25	831.61	39.83	68.82	11.00	155.33
MG26	342.13	40.01	38.67	10.67	20.00
MG27	1320.17	40.34	76.28	12.10	269.71
MG28	960.38	44.02	79.55	11.10	184.33
MG29	637.50	42.85	70.60	10.90	134.24
MG3	399.21	43.17	60.10	10.00	94.33
MG30	665.67	42.71	77.92	10.67	157.33
MG4	463.82	42.37	65.04	9.86	82.19
MG5	116.60	39.40	43.51	9.95	19.05
MG6	494.58	42.27	63.16	10.00	83.71
MG7	175.13	39.84	48.74	10.33	31.86
MG8	633.33	42.92	68.61	11.00	106.20
MG9	134.59	38.66	44.54	10.38	38.38
Means	498.8	42.0	63.0	10.6	100.4
SE ±	45.6	0.3	2.1	0.1	9.4

In this study principle component (PC) 1 which explains 68.5 % of the percentage variation was associated with Number of bolls and SCY, with a latent loading score 0.52 and 0.51 respectively (Table 4). Implying that where resources are limiting, genotypic selection can only be employed based on number of bolls and SCY.

Figure 1: Screen plot eigen values and roots (Principle components). Root (PC) 1 and 2 represented with eigenvalue 3.5 and 2 respectively

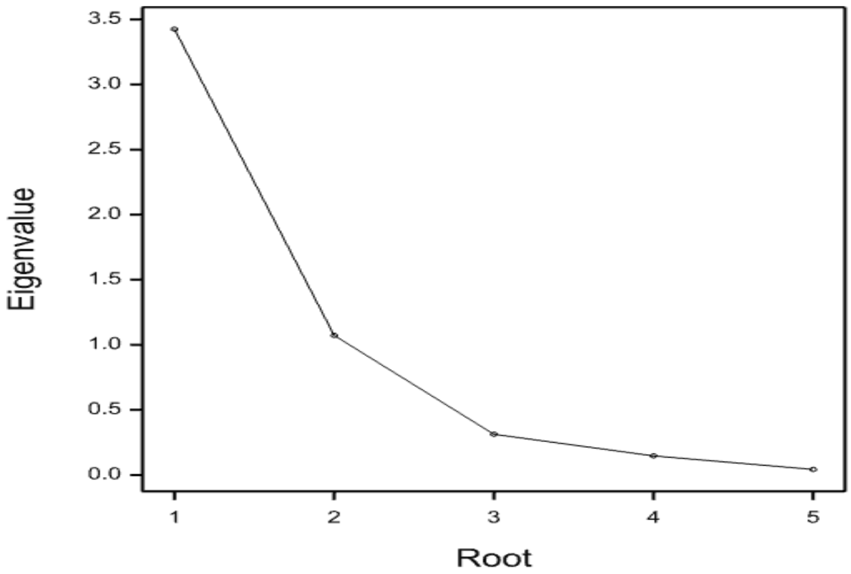


Table 4: Latent loadings of the measured parameters corresponding for computed principle components 1 to 5

Parameters	**1** (68.5%)	**2** (21.5 %)	**3** (6.3%)	**4** (2.9%)	**5**(0.9%)
Ginning out turn	0.29	0.77	0.42	0.37	0.12
Number of bolls	0.52	-0.13	-0.38	0.06	0.75
Plant height	0.49	0.29	-0.11	-0.76	-0.28
Seed index	0.40	-0.53	0.75	-0.07	0.02
Seed cotton yield	0.51	-0.18	-0.32	0.52	-58

Genotypic grouping

The scatter plot (Figure 2) revealed that the genotypes clustered into five distinct sets of which MG27 and MG 25 and MG 28 were singletons. Group C consisted of, MG5, MG7, MG9, MG10, MG26 and D the largest group, consisted of MG1, MG2, MG3, MG4, MG6, MG12, MG13, MG17, MG18, MG19, MG20, MG21, MG22, MG23, MG24, MG29, MG30. Further analysis, using cluster revealed that these groups were generated at similarity level of 97.5 % (Figure 3).

Detailed analysis using the similarity matrix (Data not shown) revealed that genotypic pair MG27 and MG5 were most dissimilar genotypes with a similarity score of 29.7 %.

Figure 2: Scatter plot for Principal Component analysis with a total percentage variation explained of 90 %. Two cluster groups C and D and singletons A, B and C were generated giving a total of five

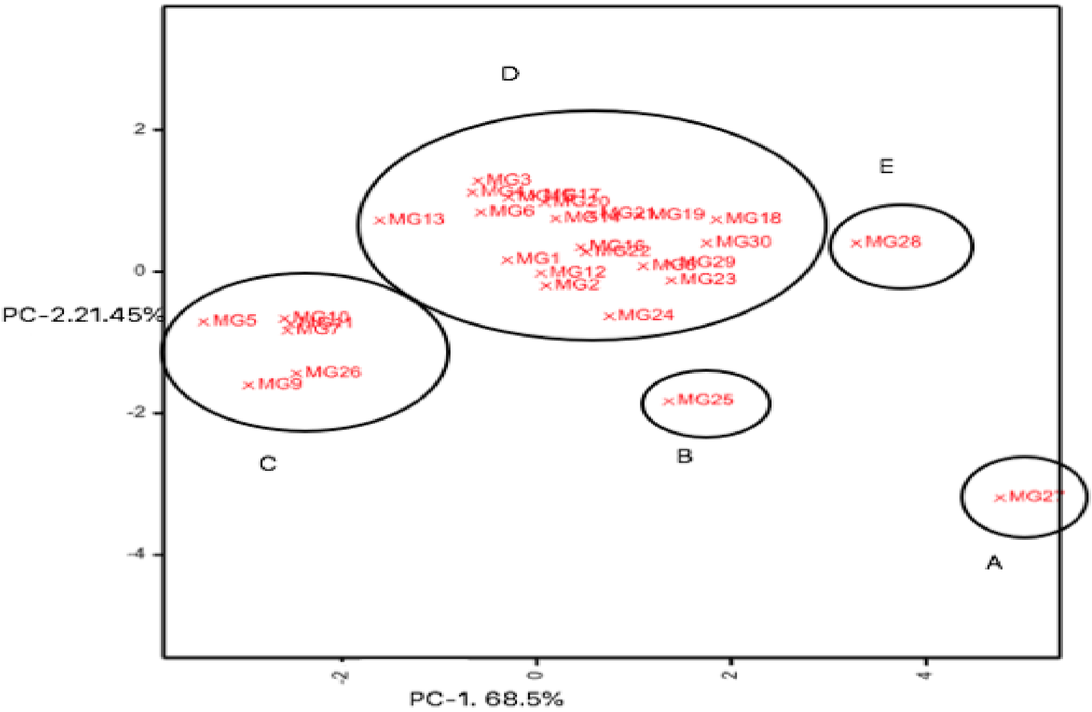
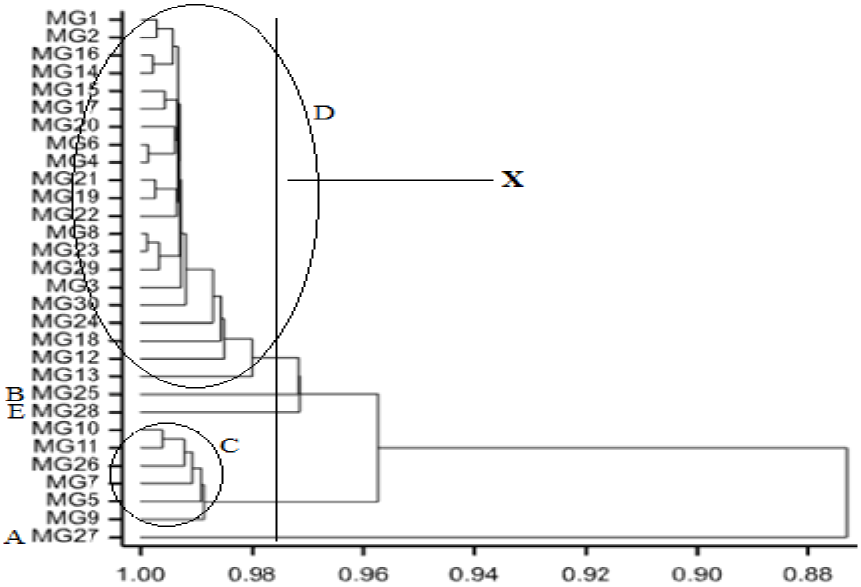


Figure 3: Dendrogram showing genotypic groupings as generated by cluster analysis. X- Line at 97.5 % cluster sets. This line crosses five lines depicting the tail end of each set. Two cluster groups C and D and singletons A, B and C were generated



Comprehending the genetic relationships among germplasm is particularly useful in breeding programs. Such information can be used in planning crosses, assigning heterotic groups, and in precise identification with respect to plant varieties (Sigh and Gupta, 2019). Principle component and cluster analysis have been used as vital analytical multivariate tools. In this study a two-dimensional PCA scatter grouping and cluster analysis (at 97.5% similarity) generated two cluster groups and three singletons (Figure 2 and 3). The reliability of the generated cluster groups from the two-dimensional PCA scatter is likely to be associated with total percentage variation explained (Chiseche *et al.,* 2020). The higher the combined total phenotypic percentage variation of the two PC scores, the more reliable the two-dimensional scatter plot is expected to be. In this study a combined PC1 and PC2 gave an approximate higher value of 90 %. The fact that the cluster pattern obtained in a PCA analysis (Figure 2) was replicated in cluster analysis (Figure 3) entails that the results were reliable. In this study five cluster sets, A, B, C, D, E were generated at a similarity index of 97.5 % as earlier mentioned. Clustering at such a level is expected among genotypes within the same species. The clustering of the genotypes at a higher level could be due to selection over time, which may ultimately have led to concentration of the elite lines within a similar gene pool (Esbroeck *et al.*,1998). Lower percentage similarity level clustering is common among different species or genera and is a common feature occurrence in evolution studies (Gori *et al.,* 2016). From this study we can deduce that selection of parents to utilise in generating hybrids or in creating of a variability of offsprings for further breeding, should come from two distinct sets. It was suggested that effective generation of diverse offspring to select from and creation of a molecular mapping population should employ a careful selection of two diverse genotypes, especially for traits which are quantitatively inherited (Tembo *et al.*, 2014; Acquaah, 2007). Chapepa *et al.*, 2020 used the same approach to identify 3 clusters of morphological traits for verticillium wilt disease variation in Cotton in Zimbabwe. In this study the least similar paired parental genotypes were identified as MG27 (from cluster set A) and MG5 (From cluster set C) with a similarity value of 29.7 %. This parental combination is expected to create maximum genetic variability among offspring for further selection. It is expected that the F1 product (MG27 X MG5), when advanced, could create maximum genetic variability among offspring generating a wider spread of choice in selecting for desirable genotypes for release or being used as parents in other crosses.

Conclusion

In this study, the parameters Number of Bolls and SCY were identified as the best at discriminating genotypic performance responses. Furthermore, the cotton genotypes clustered into five distinct sets. The most dissimilar paired parental genotypes were identified as MG27 (from cluster group A) and MG5 (From cluster group C) with a similarity value 29.7 %. This parental combination is expected to create maximum genetic variability among offspring, generating a wider spread of choice in selecting for desirable genotypes for release or being used as parents in other crosses.

Acknowledgements

The authors acknowledge and are grateful for the financial support from the Cotton Development Trust, Zambia for funding the research.

Financial support

This work was sponsored by the Cotton Development Trust based in Mazabuka Zambia.

References

1. Acquaah,,G. 2007.Principles of plant genetics and breeding. Blackwell Publishing Ltd Department of Agriculture and Natural Resources, Langston University, USA.
2. Asha, R., Lal Ahamed, M., Ratna, B.D. and Anil, K. P. 2013.Character association and path coefficient analysis for yield and component traits in upland cotton. J. Cotton Res.,29:31-35.
3. Chapepa, B., Mapuranga, R.. and Manjeru, P. 2020. Characterisation of upland cotton grown in Zimbabwe using agronomical and morphological markers for Verticillium wilt tolerance. Cogent Food Agric., 6: 766734
4. Chiseche, M., Mataa,M., and Tembo,L. 2020. Nature of Inheritance to Heat Stress and Selection of Inherent Genotypes in Tropical Maize. Asian J. Crop Sci., 5: 40-47
5. Cotton Development Trust. 2015.Cotton News. Mazabuka, Zambia.
6. Dhivya, R., Amalabalu, P., Pushpa, R. and Kavithamani, D. 2013.Variability, heritability and genetic advance in upland cotton (*Gossypium hirsutum* L.). Afr. J. Plant Sci., 8:1-5.
7. Egbuta, M.A., McIntosh, S., Daniel, W.L.E., Vancov, T. and Liu, L. 2017.Biological Importance of Cotton By-Products Relative to Chemical Constituents of the Cotton Plant. MDPI, Basel, Switzer.,22:1-25
8. Esbroeck, V., Bowman, G.A.D.T., Calhoun, D.S. and May, O.L. 1998.Changes in the genetic diversity in cotton in the USA from 1970 to 1995.Crop Sci.,38: 33-37.
9. Evgenidis, G., Traka-Mavrona, E. and Koutsika-Sotiriou, M. 2011.Principal Component and Cluster Analysis as a Tool in the Assessment of Tomato Hybrids and Cultivars. Int. J. Agron., 697879:1-7
10. FAO. 2018.Economic importance of Cotton in Burkina Faso.Background paper to the UNCTAD-FAO.
11. Gori, K., Suchan, T., Goldman, N. and Dessimoz, C. 2016.Clustering genes of common evolutionary history. Oxford University Press,UK., 33: 1590–1605.
12. ICAC. 2019.Cotton Production and Trade Trends. Cotton Report. Washington DC, USA.
13. Khan,M., Mohibullah,M., Shah,S.N.M., Amin,M. and Asmat.S. 2018. Genetic variability and morphological characterization of the elite upland cotton germplasm (*Gossipium hirsutum* L.). J. Genet. Genom. Plant Breed., 2: 1-8.
14. Mbwando, A., Lungu, D.M., Tryphone, G.M. and Tembo, L. 2016. Nature of resistance of cowpea *Alectra vogelii* infestation. Afr. Crop Sci. J., 24: 389 – 395.
15. Murtaza, N., Kitaoka, M. and Ali, G.M. 2005.Genetic differentiation of cotton cultivars by polyacrylamide gel electrophoresis. J. Cent. Eur. Agric., 6: 69-76.
16. Payne, R.W., Murray, D.A., Harding, R.A., Baird, D.B. and Soutars, D.M. 2010.An introduction to GenStat for Windows VSN International.13th Edn: Hemel Hempst, UK.
17. Simasiku, M.L., Lungu, D.M., and Tembo, L. 2020. Genotype by environment interaction of cotton, genotypes for seed cotton yield in Zambia. Asian J. Crop Sci., 5: 20-28.
18. Singh, S. and Gupta, S.K. 2019.Formation of heterotic pools and understanding relationship between molecular divergence and heterosis in pearl millet. [*Pennisetum glaucum* (L.) R. Br.]. PLoS One. 14: e0207463.
19. Tembo, L., Asea, G., Gibson, P.T. and Okori, P. 2014. Quantitative Trait Loci for Resistance to Stenocarpella maydis and *Fusarium graminearum* Cob Rots in Tropical Maize. J. Crop Improv., 28: 214-228.
20. Tembo, L. and Munyinda, K. 2015. Clustering common bean mutants based on heterotic groupings. Afr. Crop Sci. J., 23:1-7.