IJAMT

The International Journal of Agriculture, Management and Technology, Vol 4, No. 2, 2020

ISBN: 978-978-54468-2-7

www.ijamt.com.ng

HIERARCHICAL CLUSTERNG OF CASSAVA GENOTYPES BASED ON DISEASE RESISTANCE AND GROWTH COMPONENTS.

Nnabue, I., Okereke, N.R., Chukwuemeka, O.S., Kelechukwu, E.S.,Aroh, K.O., Njoku, D.N.
National Root Crops Research Institute Umudike, Abia State, Nigeria.
Phone: +2348032410397; Email: ikennannabue@yahoo.com

ABSTRACT

The knowledge of phenotypic variation in cassava (Manihot esculenta) genotypes helps in the selections of contrasting genotypes for breeding programs. This study was aimed at defining the homogeneous groups of cassava genotypes based on disease traits such as cassava mosaic disease (CMD), cassava bacteria blight (CBB) and cassava anthracnose disease (CAD). The growth component traits are canopy diameter, stem height and number of stems. The grouping was done using complete-linkage hierarchical clustering algorithm. In general, five genotypes belonging to Group 4were highly resistant to CMD, CBB and CAD.In contrast, genotypes in Group 2 and 3showed good agronomic traits such as higher canopy diameter, plant height and number of stems.Complete-linkage clustering algorithm allowed the formation of consistent groups based on growth components and disease resistance, hence was efficient in classification of cassava genotypes especially when it concerns a small amount of data.

Key words: Cassava genotypes, Cassava diseases, Clustering algorithm

INTRODUCTION

Cassava (*Manihot esculenta***)** is grown in developing countries of Africa, Asia and Latin America. The crop is mostly cultivated by the poorer sections of the population as an affordable, nutritional staple food. According to Howeler, *et al*., (2013), cassava is known for high tolerance to drought, therefore it has been considered as one of the best alternatives for providing food for the world population in the context of climatic change. There are around 10,000 varieties in this tuber crop with different characteristics (Andres and Dimuth, 2011). The production of cassava roots was calculated around 172.4 million metric tons in developing countries in 1993, and it is estimated that the yield will reach around 290.3 million metric tons by 2020, which value is worth about 13,937 million USS (Scott *et al*., 2000). However, cassava is vulnerable to a broad range of diseases caused by viruses, bacteria and fungi. Among them are Cassava Mosiac Disease (CMD), cassava bacteria blight (CBB) and cassava anthropod disease (CAD) which has significantly affected its production.This has led to the growing interest in the investigation of disease control strategies in many agricultural research programs such as breeding for resistant varieties which appears to be the most efficient means of controlling CAD, CMD and CBB (Muimba, 1982).A major part of the breeding and selection work at the Tuber and Root Crop Improvement Programme (TRIP), International Institute of Tropical Agriculture (IITA), Nigeria is on the development of resistant lines to major economic diseases: anthracnose, bacterial blight and mosaic virus(Fokunang*et al*., 2001).Genetic improvement of cassava has achieved significant progress, especially increasing the crop yield potential through the development of new varieties (De Oliveira, 2016).

As regards data analysis, there are several methods available in analyzing phenotypic variations of cassava genotypes which include clustering techniques (Hierarchical and non-hierarchical). According to Ronning*et al*., (2003), non-hierarchical clustering algorithms have emerged as better algorithm especially when there is large data.Cluster analysis as one of many multivariate statistical techniques by hierarchical methods or otherwise, classifying objects into groups (Ferraudo, 2010) with high internal consistency (within groups) and high external heterogeneity (between groups) (Hair*et al*., 2009).Clustering analysis is an unsupervised classification method, and by classifying data into different clusters, we can observe the data characteristics in different clusters through clustering algorithm without relying on pre-defined training examples. Classification and clustering have become an increasingly popular method of multivariate analysis over the past two decades, and with it has come a vast amount of published material. Different clustering algorithms have been proposed in accordance with the problems in different areas (Xu and Wunsch, 2005). It has also been used by some researchers as a means of representing the structure

of data via the construction of dendrograms (Hartigan, 1967) or overlapping clusters (Arabie*et al*., 1981; Shepard and Arabie, 1979). In discriminant analysis and automatic interaction detection, clustering makes no prior assumptions about important differences within a population. Previous studies have revealed that clustering analysis can effectively classify spectra of crop pests (Ji'An *et al*., 2018). In agglomerative clustering, single-linkage and complete-linkage have emerged as the best clustering algorithm for classification of the diseases. (Sampson, 2017).

Therefore, the specific objective of this study was to identify the optimal number of groups (clusters) and define homogeneous groups of cassava genotypes based on disease resistance and growth components traits.

MATERIALS AND METHODS

Study Area

The study area is National Root Crop Research Institute (NRCRI), Umudike in Abia State, Nigeria.The institute is strategically located at about 8 km South east of Umuahia (Abia State capital) along the Umuahia-IkotEkpene Federal Road. By road,it is 130km North of Port-Harcourt sea port, 135km South of Enugu International Airport and less than 70km East of Owerri Cargo Airport. It is situated on latitude $5^1/_2$°N and on longitude $7^1/_2$°E and 122m above sea level.

Data

The cultivars were obtained from the Genetic Resource Unit of institute. The data were obtained from Eastern field of National Root Crops Research Institute Umudike. The data indicated thirty (30) newly developed cassava genotypes. The diseasecounted and scored were severity and incidence of cassava mosaic disease (CMD), cassava bacteria blight (CBB) and cassava anthracnose disease (CAD).The severity score of 1-5 as described by Lokko*et al*., (2007) was used. The severity score of 1 indicates no disease symptom (highly resistant); class [1.1-2] indicatesmild symptoms of disease (resistant); class [2.1-3] indicatesmoderately severe disease (susceptible);class [3.1-5] indicates highly severe symptoms of disease (highly susceptible). The growth components measured were plant height (Ht) measured inmeters (m), canopy diameter (CAD) measured in meters (m), and number of stems (nStem) as a count variable.

Experimental Design

The field experiment was installed in 2016 at the eastern farm of NRCRI Umudike. The experiment was conducted in a randomized complete block design (RCBD) with thirty (30) cassava genotypesreplicated three (3) times. Cultivation was performed according to crop`s recommendation with 1m ridge spacing and 1m spacing between crops. At nine months after planting, the evaluation of the disease symptoms of cassava mosaic disease(CMD), cassava bacteria blight (CBB) and cassava anthracnose disease (CAD) were performed under field conditions. Growth components such as plant height (Ht), canopy diameter (CAN) and number of stems (nStem) were also measured and recorded.

Data analysis

The Elbow statistical method known as the scree-plot was used to determine the optimal number of groups for the thirty(30) cassava genotypes. The basic idea was to observe stability of within-group sum of squares where increase in the number of groups contributes little to the decrease in the total within group sum of square.The total within-cluster sum of square (wss) measures the compactness of the clustering and we want it to be as small as possible (Kaufman and Peter, 1990). The within sum of square to be minimized is given by:

$\text{minimize}\left(\sum_{k=1}^{k} W(C_k)\right)$(Charrad *et al*.,2014). Where C_k is the K_{th} cluster and $W(C_k)$ is the within-cluster variation.

The data was subjected to hierarchical clustering with emphasis on complete-linkage clustering algorithm. The complete linkage method ensures that all variables in a group (cluster) are within some maximum distance to each other. Thus, If AB is the distance between their furthest variables, then,

$d_{AB} = \max_{i \in A, j \in B} d_{ij}$

where d_{AB} is the maximum distance between the paired variables AB with another group (cluster).When D_{ij} is the greatest of the distance between element of *i* and element *j*then

$D_{kij} = \frac{1}{2}\left[D_{ki} + D_{kj} + \left|D_{ki} - D_{kj}\right|\right]$ (Sampson, 2017).

The analysis was done on the open source R environment version 3.6.2 (R "Dark and Stormy Night" 2019).

RESULTS AND DISCUSSION

The identification of the optimal number of groups was based on the stability of within-group sum of squares whose increase in the number of groups contributes little to the decrease in the sum of square. According to Table 1 and Figure 1, four (4) groups were identified as the optimal number of groups (clusters) to represent diseases and growth components of the cassava genotypes.

Table 1: Total sum of square for different number of clusters

No of Groups	Total WSS	No of Groups	Total WSS
1	290	11	39.9
2	234.8	12	36.1
3	193.3	13	27.4
4	166	14	27.2
5	147.8	15	22.6
6	115.3	16	18
7	94	17	15.7
8	66.4	18	23.1
9	58.2	19	14.1
10	51.2	20	11.1

WSS: within sum of square

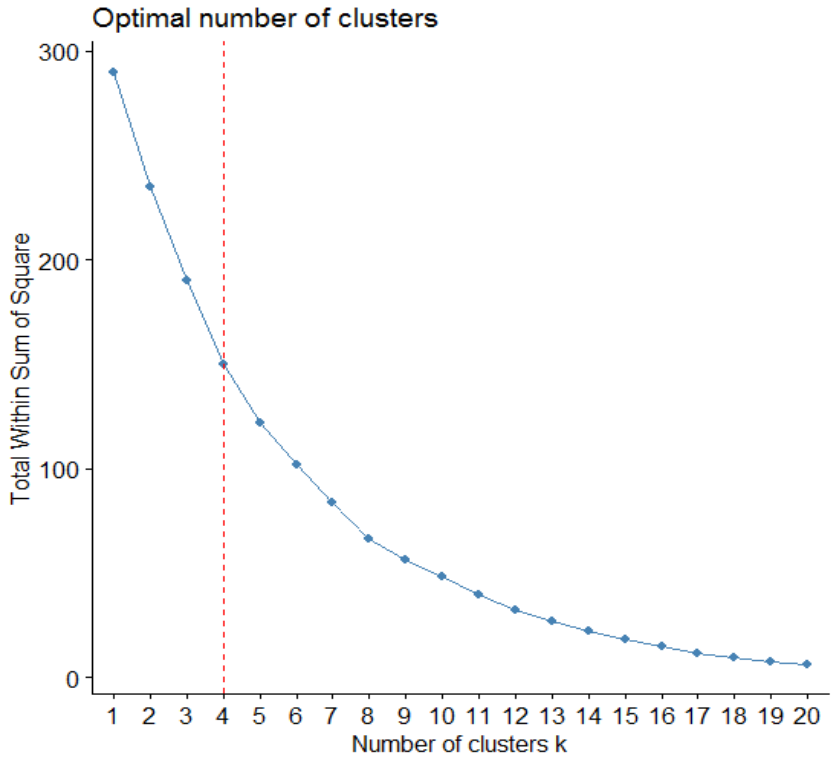

Figure 1: display of number of Groups against total within sum of squares

Hierarchical cluster analysis was carried out on disease count and growth componentprofiles of thirty (30) cassava genotype data. A complete breakdown of genotypes classifications based on diseases and growth components of these genotypes is provided in Figure 1. Four segments (clusters) were identified with **Group 1** having eight genotypes (COB-4-27, NR060333, NR110176, NR110228, NR110490, CR528-26, NR151018, NR100325), **Group 2** seven genotypes (NR100449, NR090142, NR060251, COB-5-01, MM96JW1, CR44-6, NR110376), **Group 3** with ten genotypes (NR090176, NR100012, NR100433, NR110223, NR110372, NR110512, NR090182, TMS051740, NR100417, CW451-33) and **Group 4** with five genotypes (TMS011797, TMS011097, NR110179, NR110257, NR1S10046).

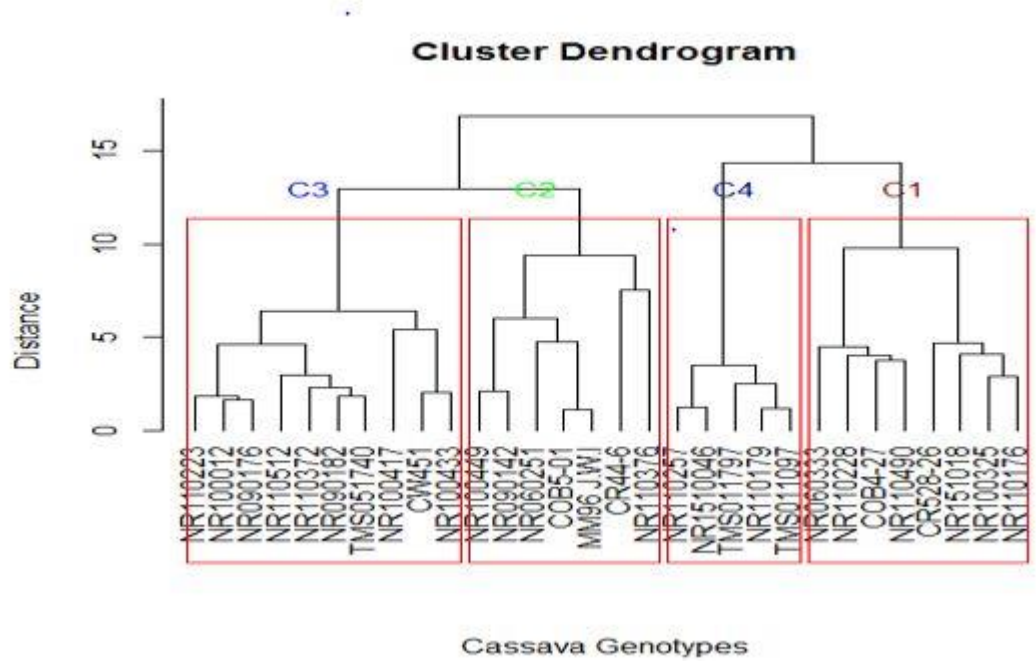

Figure 2: Cassava genotypes classification based on disease resistance and growth components profile by Complete-linkage method. Coloured rectangles indicates groups.

The choice of this number of groups (clusters) showed a high genotypic variability in the genotypes whose mean values are presented in Table 2. The number of genotypes in each group ranged from 5 (Group 4) to 10 (Group 3), averaging 7.5genotypes per group**.** Group 1 (C1) has the highest mean valueof severity and incidence of cassava mosaic disease CMD and cassava anthracnose disease (CAD) Table 2. Cassava genotypes in Group 2 showed ahigher average score of severity and incidence ofcassava bacteria blight (CBB). Group 4 (C4) contains genotypes with mean severity score 1.00 for CMD, CBB and CAD; these genotypes were considered highly resistance. In brief, the genotypes TMS011797, TMS011097, NR110179, NR110257, NR1S10046 are highly resistant to CMD, CBB and CAD. Therefore, C4constitutes promising group to select genotypes for breeding programs for disease resistance.This result is similar to those of Dallastra*et al*., (2014)who usedK means clustering algorithm in soybean breeding program to select F3 progenies by their ability to link progeny according to the most important characteristics in each cluster. The result also collaborate the work of Hounge, *et al*., (2019) who used agglomerative hierarchical clustering to select cassava mosaic disease resistant African cassava cultivars.Group3contains genotypes with mean severity score 1.00 for CMD and CBB and can also be characterized as highly resistance to CMD and CBB only.Cassava genotypes in Group2 has the highestmean value for canopy diameter and while genotypes in Group3 are characterized with highermean scores for plant height and number of stems. These two groups also constitute promising groups to select genotypes for breeding programs like breeding for commercial use. This result is similar to those of De Oliveira *et al*. (2016) who used K-means clustering algorithm to identify groups showing highest yield for noncommercial roots (YiNComRoo) of cassava accessions.

Table 2: Average mean values of the groups based on the diseases and growth components

Groups	CMDi	CMDs	CBBi	CBBs	CADi	CADs	Ht	nStem	CAN
C1	2.00	1.75	0.00	1.00	2.00	2.38	1.27	1.67	1.61
C2	0.29	1.14	0.71	1.14	0.43	1.29	1.60	1.57	2.43
C3	0.00	1.00	0.00	1.00	0.40	1.40	1.73	1.91	1.34
C4	0.00	1.00	0.00	1.00	0.00	1.00	1.18	1.18	1.43

CMD: cassava mosaic disease; CBB: cassava bacteria blight; CAD: cassava anthracnose disease; Ht: plant height; CAN: canopy diameter; nStem: number of stems; i= incidence; s= disease severity.

CONCLUSION

In the course of this research, 30 newly developed cassava genotypes were used for classification. Of the thirty (30) genotypes scored against cassava mosaic disease (CMD), cassava bacteria blight (CBB), cassava anthracnose disease (CAD), plant height, number of stems and canopy diameter, five (5) genotypes; TMS011797, TMS011097, NR110179, NR110257 and NR1S10046were morphologically resistant to CMD, CBB and CAD. These genotypes may be prioritized for breeding programs to improve CMD, CBB and CAD resistance in other cassava genotypes. Similarly, genotypes in Group 3 (C3);NR090176, NR100012, NR100433, NR110223, NR110372, NR110512, NR090182, TMS051740, NR100417 and CW451-33 showed the higher score for plant height and number of stems. These genotypes identified in this work can help farmers make selections ultimately for seed propagation.

REFERENCES

Andres, S., &Dimuth, S. (2011). Insights into the physiological, biochemical and molecular basis of postharvest deterioration in cassava (*Manihot esculenta*) roots. *American Journal of Experimental Agriculture1(4): 414-431*.

Arabie, P., Carroll, J.D., DeSarbo, W. & Wind, J. (1981). Overlapping Clustering: A New Method for Product Positioning. *Journal of Marketing Research.*Volume: 18 issue: 3, page(s): 310-317

Charrad M., Ghazzali N., Boiteau, V. &Niknafs A. (2014). NbClust Package for Determining the Best Number of Clusters. *Journal of statistical software* 61(6):1-36. DOI: 10.18637/jss.v061.i06

Dallastra, A., Unêda-Trevisoli, S.H., Ferraudo, A.S. & Di Mauro, A.O. (2010). Multivariate approach in the selection of superior soybean progeny which carry the RR gene. *RevistaCiênciaAgronômica*, v. 45, n. 3, p. 588-597.

De Oliveira, E.J., Fabiana F.A., , Cinara Fernanda Garcia Morales, C.F.G., , Saulo Alves Santos de Oliveira, S.A.S. & Santos, V.S. (2016). Non-hierarchical clustering of Manihot esculenta Crantz germplasm based on quantitative traits. *RevistaCiênciaAgronômica*, v. 47, n. 3, p. 548-555

Ferraudo, A. S. (2010). Multivariate analysis techniques: An Introduction. Sao Caetano: StatSoft South America

Fokunang, C.N., Dixon, A.G.O., Ikotun, T., Tembe, E.A., Akem, C.N. &Asiedu, R. (2001). Anthracnose: An Economic Disease of Cassava in Africa. *Pakistan Journal of Biological Sciences, 4: 920-925*

Hair, J.F., Sant'Anna, A.S. &Gouvêa. M.A. (2009). Multivariate data analysis. six edition. Porto Alegre Bookman. P. 688

Hartigan, J.A. (1967), Representation of Similarity Matrices by Trees, *Journal of the American Statistical Association, 62*, 1140–1158.

Houngue, J.A., Zandjanakou-Tachin, M., Ngalle, H.B.,Pita, J.S.,Cacaï, G.H.T., Ngatat, S.E., Bell, J.M., &Ahanhanzo, C. (2019).Evaluation of resistance to cassava mosaic disease in selected African cassava cultivars using combined molecular and greenhouse grafting tools. *Physiol Mol Plant Pathol*. 105: 47–53.doi: 10.1016/j.pmpp.2018.07.003

Howeler R. L., Thomas N., Holst Sanjuán K., Sanjuán K. H., Quirós H., Isebrands J. G., et al. (2013). Save and Grow: Cassava. A guide to Sustainable Production Intensification Produire plus Avec MoinsAhorrar Para crecer. Roma: FAO.

Ji'An Xia., YuWang Y., HongXin C., YaQi K., DaoKuo G., WenYu Z., SiJun G. &GuangWei C. (2018). Performance analysis of clustering method based on crop pest spectrum. *Engineering in Agriculture, Environment and Food.* www.elsevier.com/locate/eaef

Kaufman, L. and Peter R. (1990). Finding Groups in Data: An Introduction to Cluster Analysis.Print ISBN:9780471878766 |Online ISBN:9780470316801 |DOI:10.1002/9780470316801

Lokko Y., Okogbenin, E., Mba, C., Dixon, A., Raji, A. &Fregene, M. (2007). Cassava In: Genome Mapping and Molecular Breeding in Plants, Pulses, Sugar and Tuber Crops. Kole C., editor. Springer-Verlag Berlin Heidelberg; pp 249-269

Muimba, K.A. (1982). Predisposition of cassava plants to infection by *Colletotrichum manihotis* Henn, and some factors involved in the initiation of anthracnose disease. M.Phil. Thesis, University of Ibadan, Nigeria. pp.241

Ronning, et al. (2003). Comparative analyses of potato expressed sequence tag libraries. *Plant Physiology*, v. 131 n. 2: 419-429. doi: 10.1104/pp.013581

Samson A.,&Mataimaki B.J. (2017). Classification of Some Seasonal Diseases: A Hierarchical Clustering Approach. *Biomedical Statistics and Informatics*2017; 2(5): 122-127http://www.sciencepublishinggroup.com/j/bsidoi: 10.11648/j.bsi.20170205.11

Scott, G.J., Rosegrant, M.W., & Ringler, M.W. (2000). Roots and tubers for the 21st century: trends, projections and policy options.*Food, Agriculture and the*

Environment Discussion Paper 31. International Food Policy Research Institute

Shepard, R. N., &Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86*(2), 87–123. https://doi.org/10.1037/0033-295X.86.2.87

Xu, R. & Wunsch D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks* .Volume: 16 , Issue: 3. 645 – 678. doi: 10.1109/TNN.2005.845141