Agro-Science Journal of Tropical Agriculture, Food, Environment and Extension
Volume 7 Number 1 January, 2008 pp. 66 -71
ISSN 1119-7455

CHARACTERIZATION OF MAIZE *(Zea mays L.)* GERMPLASM WITH PRINCIPAL COMPONENT ANALYSIS.

Okporie E. O.
Department of Crop Production and Landscape Management, Ebonyi State University, Abakaliki.

ABSTRACT

Seventy-one (71) open pollinated maize genotypes of diverse agronomic and chemical characteristics were obtained from the germplasm collection of the University of Nigeria, Nsukka for this study. The genotypes were planted out in an ear-to-row fashion in April 2000. In order to have a well organized and streamlined maize research program, it became important that these genotypes be characterized according to similarities in traits. Thus, the genotypes were subjected to principal component analysis; a tool which has the ability of grouping genotypes of identical genetic characteristics. Following the analysis, seven components whose eigenvalues were equal or more than one, accounted for more than 75% of the total variation in the data and centered on plant structure, yield, quality (lysine, protein, oil omylose and amylopectin) and maturity date. Consequently, the entire genotypes were successfully grouped into two, early and late maturing composite varieties with high yielding and high protein characteristics. The early maturing composite variety had mean number of days to 50% silking of 58 days and comprised 18 genotypes while the late maturing composite variety had mean number of days to 50% silking of 61 days and comprised 53 genotypes.

Key Words: *principal component; Zea mays; germplasm; analysis; characterization*

INTRODUCTION

Maize has been described as a golden crop because every part of the crop is useful to man and his animals. It is a cross-pollinated crop and like any other cross pollinated crop, the phenotypic and genotypic attributes are controlled by qualitative and quantitative genes and thus, can be altered through genetic manipulations and selections. According to Winter (1974) and Moll and Stuber (1974), the effectiveness of manipulation and selection of desirable alleles rely mostly on the amount of diversity and the extent of heritability of the desirable alleles of the germplasm that form the base population. Equally important is the total number of the germplasm because the greater the number in a collection, the greater the chances of accumulating desirable heritable characters.

Principal component (PC) analysis is a statistical tool which helps in determining components that account for variation in a reduced data set derived from a large number of variables in the data. It is used to find the best linear combinations of variables that would account for more of the variances in the data as a whole than any other linear combination of variables. Therefore, the first principal component (PCI) may be viewed as the single best summary of linear relationships exhibited in the data. The second principal component (PC2) is defined as the next best linear combination of variables under the condition that the second component is orthogonal (unrelated) to the first component, the second must account for the proportion of the variance not accounted for by the first one. Thus, the second principal component may be defined as the linear combination of variables that account for the most residual variance after the effect of the first component is removed from the data. Subsequent components are defined similarly until all the variances in the data are exhausted. Unless at least one variable is perfectly determined by the rest of the variables in the data, the principal component solution requires as many components as there are variables (Norman *et al.* 1975). However, to determine the number of components that are to be considered meaningful and retained for an interpretation, there are two criteria that can be used. The first requires an examination of the

eigenvalue. A principal component whose eigenvalue (Latent root) is equal to or greater than one (≥1) is considered important and worthy of interpretation. The second criteria is to examine the graph of the eigenvalues against the component numbers, the point at which the resulting curve begins to flatten is cut-off as unimportant and thus not meaningful (Kaiser, 1960; Cattell, 1966; Richard *et al.* 1980; Iezzoni and Pritts 1991).

The interpretation of the output from a principal component analysis requires that one determines the importance of principal components and the variables associated with each principal component (Iezzoni and Pritts 1991). There are no rules for making this determination, unlike with unvariate statistics, so good judgment must be used. According to Iezzoni and Pritts (1991), variables with large eigenvectors or loadings either positive or negative are considered to be contributors to the principal components. But where there is no large eigenvectors or loadings, for example, where several variables are highly correlated, good judgment and biological meanings attached to a particular variable should be considered. In fact, for the interpretation of principal component analysis, good judgment and biological meaning attached to a particular variable is paramount. Thus, PCA was employed in this research to help characterize and group 71 maize genotypes according to their similarity in agronomic and chemical attributes.

MATERIALS AND METHODS

This research was conducted in the Department of Crop Science, University of Nigeria, Nsukka. Seventy one (71) open pollinated genotypes of maize were collected from the germplasm collection of University of Nigeria, Nsukka. The genotypes were given numbers while in the gene bank and were planted in ear-to-row progeny method according to their numbers on April 21, 2000, on a well pulverized flat plot. The plant spacing was 75 x 25cm at one plant per hill giving a plant population of 53,333 plants per hectare.

Desirable plants in each genotype in terms of phenotypic appearance were self-pollinated by hand using tassel-bag-shoot bag procedure (Obi, 1991). Recommended agronomic practices were performed throughout the growing period of the plants. Harvesting was done when the plants had completely senesced and had reached physiological maturity using black layer formation as an index of maturity (Baker, 1973). Some agronomic characteristics such as plant and ear heights, days to 50% tasselling and silking, tassel number per plant, ear number per plant, cob circumference, 100 – seed weight and kernel density were recorded. After harvest the ears of the genotypes were shelled and equal quantities of seeds were analysed for protein, lysine, oil, amylose, amylopectin and sugar. Percentage protein content was determined using Micro-kjeldahl method, described by Pearson (1976). Available lysine was determined using method of 2, 4, 6, Trinitrobenze-1-sulfonic Acid (TNBS) as described by Obi (1982). Proximate system for food analysis which employs the Soxhlet's extractor as described by Anon (1978) was used to determine the percentage oil content of the maize. Amylose content in the maize was determined by the method described by Hassid and McGreedy (1943). The percent amylopectin was determined by subtraction on the assumption that 100% starch = Amylose (%) + Amylopectin (%). Sugar content was determined by Gravimetric copper Reduction method as described by Pearson (1976).

RESULTS AND DISCUSSION

The seven (7) components which had eigenvalues equal to or of greater than one (≥ 1.0) were retained as meaningful and worthy of interpretation (Table 1).

These seven components accounted for 75.31% of the total variation in the data set. The principal component analysis indicated that the first Principal Component (PCI) had an eigenvalue of 2.45 and explains 16.32% of the total variation in the data set (Table 1), this suggests that PCI represents the equivalent of two individual variables and the two variables that weighted higher than the other variables are amylose and amylopectin The second component (PC2) had eigenvalue of 2.23 and accounting for 14.88% of the total variation; the value also suggests that the PC2 represents the equivalent of two individual variables which are days to 50% tasselling and silking. Principal components 3 to 7 had each more than one eigenvalue, thus they represent equivalent of one individual variable each and accounted for 11.4%, 9.6%, 8.5%, 7.7% and 6.9%, respectively of the total variation in the data set.

Table 1: Eigenvectors of the principal component (pc) axes from PC analysis of maize genotypes. Eigenvalues and their contribution to total variation are listed at the bottom of columns.

Character	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Oil	-0.2516	0.0579	-0.0708	-0.3845	0.3447	-0.049	-0.2265
Lysine		0.1910	-0.1424	-0.3115	-0.5221	0.0126	0.1457
Protein	-0.2489	-0.1349	-0.1352	-0.2866	-0.0983	0.5283	-0.0688
Amylose	-0.4779	-0.3227	0.1116	-0.3247	0.0737	-0.0537	0.0808
Amylopectin	0.4779	-0.3227	0.1116	-0.3247	-0.0737	0.0537	-0.0808
Sugar	-0007	-0675	0.4970	0.1990	0.0960	-0044	-0.3745
Plant height	0.0627	0.3779	0.4144	0.0043	0.0484	0.2726	0.3336
Ear height	0.0169	0.2486	0.4893	-0.2557	-0.0515	0.1247	0.3556
Tassel no per plant	-0.1602	-0.0972	0.0454	0.3803	-0.4761	0.3188	-0.1875
Ear no per plant	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Days to 50% Tasselling	0.3449	0.4083	-0.3394	0.0579	0.0154	0.1707	-0.0206
Days to 50% Silking	0.3769	0.3856	-0.2762	0.0766	0.1661	0.0600	-0.0937
Cob length	-0.1615	-0.1665	0.0082	0.0482	0.5470	0.2344	0.1920
Cob circumference	0.0592	-0.3238	-0.2290	0.1279	0.0913	0.4626	0.3519
100 seed weight	0.2622	-0.1097	0.1082	0.4262	-0.0205	0.0749	0.1109
Kernel density	-0.0257	-0.2343	-0.1248	0.0509	-0.0570	-0.4409	0.5607
EIGEN-VALUE	2.447	2.232	1.708	1.440	1.280	1.158	1.032
%VARIANCE	16.32	14.88	11.39	9.60	8.52	7.72	6.88

PC1 probably represents latent variables reflecting starch. PC2 could be reflecting flower development. The PC3 showed high weights in sugar, plant height and ear height, thus, the PC3 probably represents plant structure and sugar content. In the fourth principal component, (PC4) 100 - seed weight had the largest weight, thus reflecting yield. The fifth principal component showed that lysine content, cob length and tassel number per plant had the largest weight, so, PC5 probably was reflecting plant yield and lysine content. The 6th principal component, (PC6), protein, cob circumference, and kernel density weighted high, so yield and protein are probably being reflected. The seventh principal component, (PC7) reflected plant yield also.

Apart from the contributions of these variables into the component analysis, Starch (amylose and amylopectin) and lysine were the chemical contents considered most important for the grouping of the genotypes. This is because the level of lysine content in grain determines the nutritional quality of the maize. Besides protein content of maize can be improved by cultural practices (the addition of nitrogenous fertilizer, etc). Apart from the above, lysine is the first limiting amino acid in cereals for human and monogastric animals and any increase in lysine content and in protein results in improved nutritional quality (Eggum, 1977). Again, according to Newman *et al.* (1978) when lysine content in maize is increased fewer protein supplements are needed in human and animal nutrition. Starch, apart from its usefulness in food industries and other industrial products, the level of starch content in a grain suggests the level of sugar content of the grains. Also, it is said that it has been found difficult to develop high nutritional quality with high yield because of reduction of starch content of the grain (Mehta *et al.*1978; and Sen and Mehta, 1981). Oil content was not considered because all the genotypes had between low to medium oil content (Obi and Ihedigbo 1987). Maturity date (number of days to 50% tasselling and silking), plant structure (plant height and ear height); cob length and tassel number per plant were considered for agronomic characteristics. The number of days it takes a genotype to reach 50% tasselling and silking from the date of planting determines whether a genotype is early maturing or late maturing. Plant and ear heights are very important plant structure that determine how safe the plants are to handle during hybridization and measurements; how firm the plant can stand on the ground against lodging (standability) which is a very serious problem to the farmer and how suitable are the plants to mechanical harvesting. Cob length and tassel number per plant are very good parameters for predicting high yield. The longer the length of a cob of a corn if well filled with grains due to heavy availability of pollen grains the greater the number of maize grains, thus, the higher the yield. Therefore, the greater the number of tassel branches in a maize field, the more the pollens that will be produced leading to grater pollination and fertilization, thus, the greater the yield. Figures 1 - 3 are the results from the analysis of the variables mentioned above.

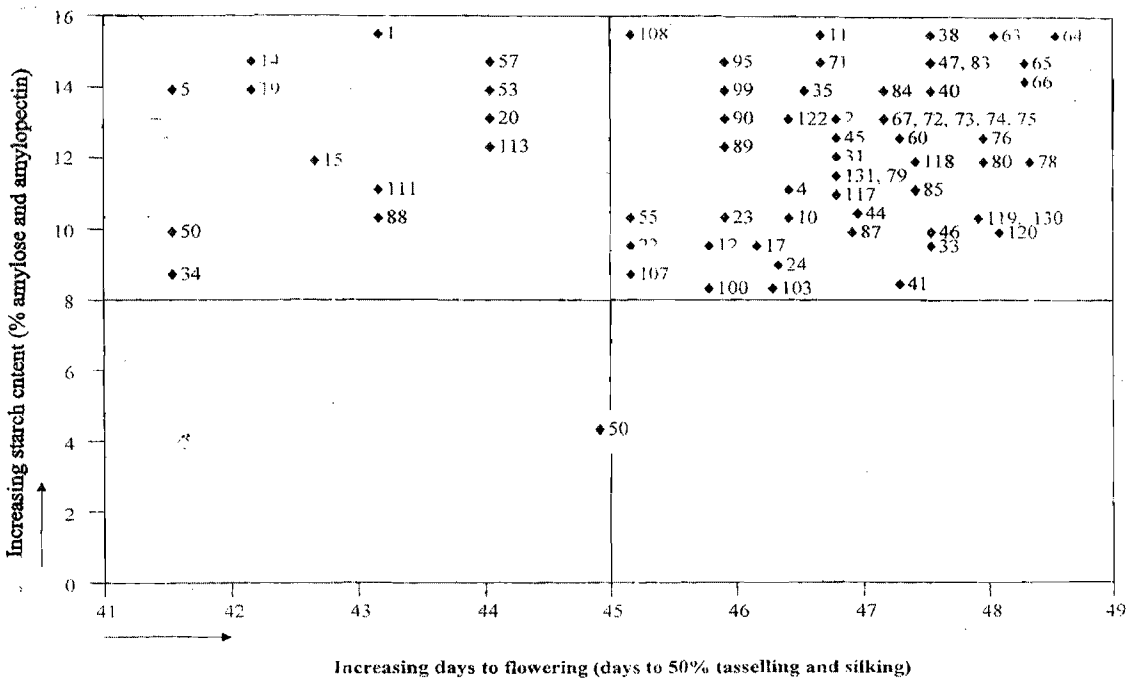

Figure 1: Plot of the first (PRIN 1) and second (PRIN 2) principal components based on yield and maurity time

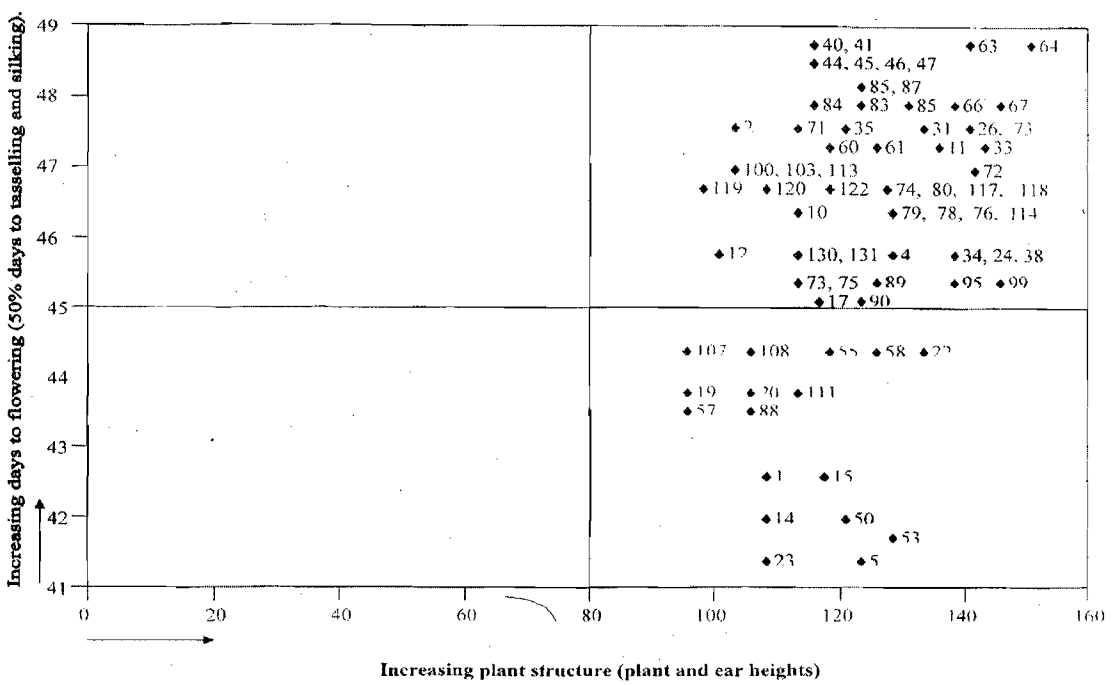

Figure 2: Plot of the second (PRIN 2) and third (PRIN 3) principal components based on maturity time and growth rate components

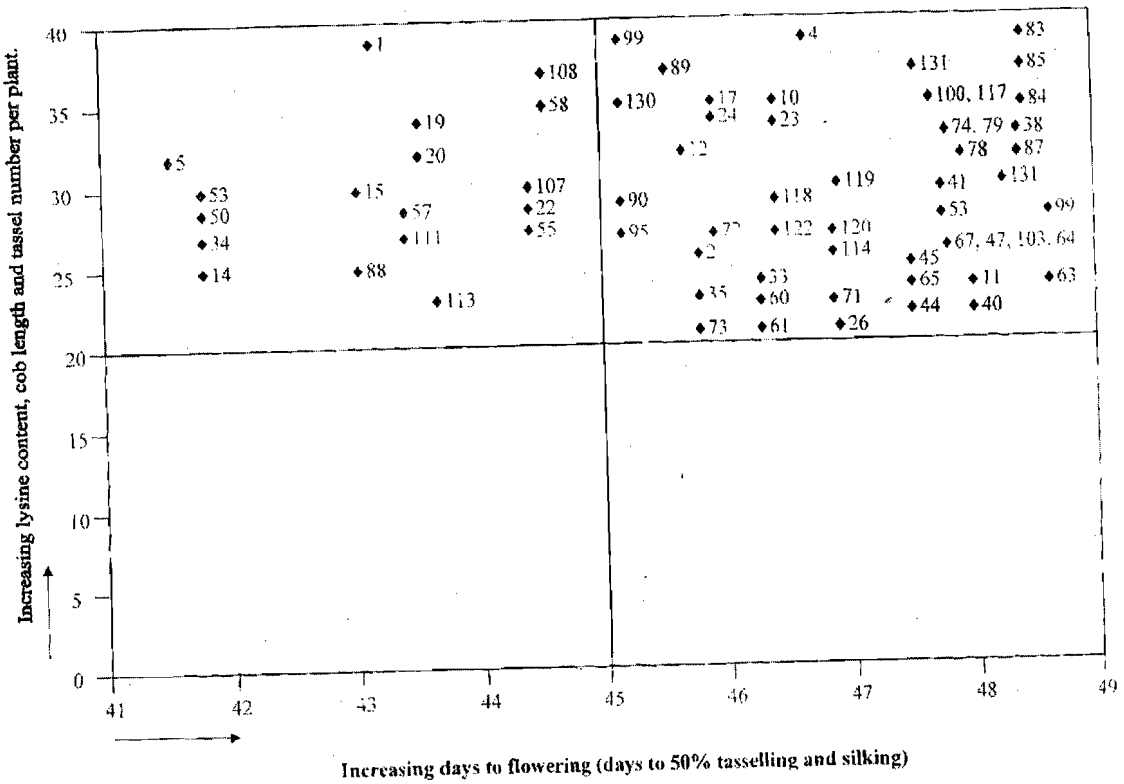

Figure 3: Plot of the Second (PRIN 2) and fifth (PRIN 5) principal components based on maturity date, chemical and yield components.

From the analysis of starch content (amylose) and days to flowering (50% tasselling and silking), the genotypes were grouped into three (3). Group 1 comprised 57 genotypes with high starch content but late maturing. The second group consisted of 13 genotypes with high starch but early maturing. The third group which had only one genotype was low in starch content (about 9% amylose content) and average in maturity date.(Figure 1). From the analysis of plant structure and flower development (Figure 2), the 71 genotypes were grouped into two. Group 1 contained early maturing genotypes while Group 2 consisted of late maturing genotypes. None of the genotypes was grouped as short in plant structure. In Figure 3, the whole genotypes were grouped into two groups. Group 1 comprised those genotypes that were early maturing, high yielding and high lysine content. Probably, this may be the most important group, because early maturity is a desirable character in plants, for farmers, just like high yield as well as high lysine which is desired in nutrition for those who use maize as their meal. The second group consisted of the genotypes with high lysine, high yield but late maturing.

These last two groupings can form two important populations useful in starting reciprocal recurrent selection and/or starting a new phenotypic recurrent selection, hybrid production or put any other genetic improvement programme. The 18 genotypes that formed a separate group from the analysis of PC2 and PC5 (Fig 3) were found to have mean number of days to 50% silking of 58 days (less than 110 days to maturity). While the fifty-three (53) genotypes, which formed the second group, had a mean number of days to 50% silking of 61 days (more than 110 days to maturity). The first group was regarded as the developed early maturing composite variety while the second group was regarded as the developed late maturing composite variety. This observation is in line with the report of Fajemisin (1985) that varieties that silked in less than 60 days from the date of planting were regarded as early maturing, while those that silked at 60 days and above were regarded as late maturing.

CONCLUSION

The seventy- one maize genotypes, with the help of principal component analysis, were successfully characterized and accurately grouped into two - Early and late maturing, high yielding and high quality maize varieties. This is useful for further maize improvement researches, to farmers and other stakeholders in maize production and usage.

REFERENCES

Anon, 1978. Determination of Oil in feeding stuffs. Department of Animal Science Laboratory Manual. University of Nigeria, Nsukka pp. 20.

Baker, R. 1973. Black –layer development. *World farming* 15: 14 –19 .

Cattell, R.B. 1966. The screen test for the number of factors. *Mulitvariate Behavourial Research* 1:243-276.

Eggum, B.O. 1977 Nutritional aspects of cereal proteins. In: Genetic Diversity in plants. (eds) Muhammed A; Aksel R. and Von Borstel, R.C. Plenum press, New York , USA pp. 349-369

Fajemisin, J. M. 1985. Status of maize production technology in Nigeria and prospects for sustained self-sufficiency. Paper presented at NAFPP, 3rd Joint workshop, Owerri, Imo State, Nigeria. March 7–9.

Hassid, W. A and McGready, R.M. 1943. Separation and quantitative estimation of amylose content of corn endosperm starch. *Agron. J.* 41:409 – 411.

Iezzoni, A.F. and Pritts, M.P. 1991 Applications of principal component Analysis to Horticultural Research. *HORTSCIENCE.* 1.26 (4): 334-338.

Kaiser, H.F. 1960, The application of electronic-computers to factor analysis. *Educational and Psychological Measurements* 20:141-151.

Mehta, S.L; Dongra, A.B., Johari, R.P., Lodha. M.L. and Naik, M.S 1978. Biochemical constraints that determine protein quality and grain yield in cereals . In: seed protein improvement in cereals and grain legumes. Proc. symp. IAEA, Vienna pp. 241-256.

Moll., R.H. And Stuber, C.W. 1974. Quantitative Genetics-Emperical Results Relevant.Plant Breeding. *Adv. Agron.* 26:277 – 313.

Newman, C.W., El-Negoumy, A.M., Eslick, R.F. 1978. Replacing soy protein in some diets with high protein barley and amino acids. *J. Animal Sci.* 46:161-166.

Norman, H.N., Hadai, C.H., Jean. G. J., Karim, S; and Dale, H. B. 1975. *Statistical package for the social sciences.* 2ed. McGraw-Hill Book Company, New York. McGraw.

Obi, I.U. 1982. Application of 2,4,6, - Trinitrobenzene – 1 – Sulphuric Acid (TNBS) Method for the determination of available lysine of maize seed. *Agric. Biol. Chem.* 46 (1): 15-20

Obi, I.U. and Ihedigbo, N.E. 1987. Amylose Amylopectin and oil contents of some Nigerian maize cultivars. *Niger. Agric. Jour*-22:91-100.

Obi I.U. 1991. *Maize, Its Agronomy, Diseases, Pests and Food Values.* Optimal Computer Solutions Ltd. Enugu, Publishers. Enugu, Nigeria. 206 pp.

Pearson, D. 1976. *The chemical analysis of foods.* Seventh ed. Churchill Livingstone Pub. Co. London and New York, 575 pp

Richard, H.L. , Peter, F.M. and Ruth, Z, C. (1980) *Introduction to Bivariate and multivariate Analysis.* Scott Foreman and Company Glenview, Illinois USA.

Sen K. and Mehta, S. L. 1981. Protein and enzyme levels during development in high lysine barley mutant. *Phyto chemistry* 1323 – 1329.

Winter, F. L. (1974). The mean and variability as affected by Continuous Selection for Composition in Corn. *Crop Sci. Soc. of Am.*:95 – 118.