Original Research Article

Genetic Diversity and Variability among Papaya (*Carica papaya* L.) Genotypes using Multivariate Analysis

ABSTRACT

Multivariate analysis was used to group and study the pattern of genetic variation and relationship among ten papaya genotypes in Benin City, Edo State, Nigeria. The experiment was conducted at the Teaching and Research Farm of the Faculty of Agriculture, University of Benin, Benin City, Edo State, Nigeria (Latitude: 6° 33′N, Longitude: 5° 37′E; 79m asl). Field evaluation of the papaya genotypes was carried out from October 2012 to June 2013. The experiment was laid out as a randomized complete block design with three replications. Euclidean genetic distance between CP006 and CP012 was the highest while the lowest Euclidean distance was between CP001 and CP011. The level of variability observed suggested a high diversity among the genotypes. The result of the principal component analysis indicated that the contribution of the first three factors with Eigen value greater than one accounted for 93.0% of the total variation. PCA and Cluster analysis produced similar results in classifying the genotypes into three heterotic groups. The first component loaded highly for fruit characters except fruit flesh thickness which loaded highly in the third component and fruit thickness which loaded moderately in the first component and thus can be labelled fruit yield. The vegetative components loaded highly in the second component and thus can be labelled as vegetative component. Thus the pawpaw genotypes can be distinguished based on either yield or vegetative characters with more reliability on the yield parameters and better resolution when both yield and vegetative component are considered together. The agronomic characters were efficient in assessing genetic divergence with Leaf width and flesh fruit thickness as the most distinguishing characters as revealed by discriminant analysis. These characteristics could be useful as markers for the selection of female parents in yield improvement programs. The three clusters formed indicates intraspecific phenotypic dissimilarity among the ten genotypes especially with the separation of the genotypes that were collected from similar environments. The phenotypic variations could be explored for utilization, conservation and for future genetic improvement by selection of genotypes with promising agronomic characters. CP012 was particularly superior with respect to the studied traits and was the only genotype in its cluster group. Molecular studies would be useful to confirm the genetic diversity and characterize these genotypes for more detailed examination.

Keywords: papaya, phenotypic variation, multivariate analysis, Benin City.

1. INTRODUCTION

The papaya (*Carica papaya* L.) belongs to the small family Caricaceae, which includes 35 species placed in six genera [1] (Ramos *et al.* 2012). Among all species, 32 are dioecious, two trioecious and one monoecious [2] (Ming *et al.* 2007). The papaya is the only species of the genus *Carica*, also being the best known and most economically important within the family [3] (Van Droogen- broeck *et al.* 2002), showing widespread cultivation in tropical and subtropical regions around the world. It is commonly known for its food and nutritional values throughout the world [4] (Krishna *et al*. 2008). Papaya is a major fruit crop worldwide that is primarily consumed as fresh fruit. It is highly abundant and is commonly known as pawpaw in Nigeria. It is an invaluable plant that is prevalent throughout Tropical African and Nigeria is the third largest producer globally [5] (FAO 2002).

Genetic diversity among parental is considered an important factor for obtaining heterotic hybrids. This diversity is one of the restraining tools for breeding programmes based on hybridization, because it generates parameters for identifying superior parental. According to [6] Cruz and Carneiro (2003), this distance is essential to increase the chance of recovering superior genotypes. Several authors have reported the efficacy of genetic divergence as a criterion for choosing parents to be crossed [7, 8, 9] (Dias and Kageyama 1997; Hamza et al Odewale *et al*. 2012). Genetic diversity plays an important role in plant breeding because hybrids between lines of diverse origin generally display greater heterosis than those between closely related strains [10] (Singh 1983) which permits to select the genetically divergent plants to obtain the desirable recombination of the segregating generation.

Multivariate analysis is a useful tool in quantifying the degree of divergence between biological population at genotypic level and to assess relative contribution of different components to the total divergence both at intra- and inter-cluster levels [11, 12] (Jatasra and Parada 1978; Zahan *et al*. 2008). The multivariate analysis such as the Principal Component Analysis, cluster and discriminant analysis have been used to uncover similarities between variables and determine the amount of variation and the most suitable combinations of genotypes for a breeding program. The objectives of this research were to study the genetic variation among ten papaya genotypes by using multivariate analysis to classify the accessions in order to identify divergent parents for breeding programme and select the most suitable combinations which would provide superior segregates, as well as, to investigate the importance of the evaluated characters.

2. MATERIAL AND METHODS

The ten genotypes used in this study (CP001, CP002, CP003, CP004, CP005, CP006, CP007, CP008, CP011 and CP012~~0~~) were obtained from Uselu market and home gardens in Benin City metropolis. The collection was undertaken in October, 2011. The seeds of each papaya accessions were sown in October, 2011 in drill rows of 4m by broadcasting and gradually thinned to three plants per row, spaced at 2m x 2m. The experiment was conducted at the Teaching and Research Farm of the Faculty of Agriculture, University of Benin, Benin City, Edo State, Nigeria (Latitude: 6° 33′N, Longitude: 5° 37′E; 79m asl). Field evaluation of the papaya genotypes was carried out from October 2012 to June 2013. The experiment was laid out as a randomized complete block design with three replications.

At flowering, about 9 months after planting, NPK 15:15:15 was applied at the rate of 0.6kg per plant. Each papaya stand received additional 0.6kg of NPK 15:15:15 at 18 months after planting. ~~Weeds were controlled manually throughout the period of the study.~~ Data were collected on stem height, stem girth, stem internodes, number of nodes, height at first fruiting, leaf length, leaf width, petiole length, fruit length, fruit diameter, fruit flesh thickness, length of peduncle, number of fruit per plant and fruit weight at about one year after planting when the first fruit was matured and ripe for harvesting.

The stem height was measured from the ground level to the stem tip with a tape, while stem girth was measured at 10th internode when counted downward from the point of attachment of the first fruit. Leaf length was obtained by measuring from the point of attachment of the petiole to the tip of the longest leaflet, while leaf width was the measured widest portion of the leaves. Petiole length was measured as the distance from the point of its attachment to the stem to the point of attachment of the palmate leaves. At harvest, the fruit was detached from the peduncle and the fruit weight was determined by weighing with a weighing scale. Longitudinal sections of the harvested fruits per tree were made, and then the fruit length was determined from pole to pole of the fruits. Fruit diameter was determined from the equator of the sectioned fruit and the flesh thickness was measured with a measuring tape.

2.1 Data analysis

The data were subjected to descriptive statistics and parameters such as mean, standard deviation (SD) and coefficient of variation (CV) for each one of the 13 studied traits were calculated. Clustering of genotypes into similar groups was performed using Ward’s hierarchical algorithm based on squared

Euclidean distances by subjecting the 10 x 13 data matrix to cluster analysis. Discriminant function analysis was used to confirm the accuracy of grouping that was produced by cluster analysis. In order to identify the patterns of phenotypic variation, principal component analysis (PCA) was conducted. The PCs with Eigen value >1.0 was considered as inherently more informative than any single variable alone [13] (Kaiser 1960).The component was further rotated using the varimax method with Kaiser Normalization. SAS [14] (SAS Institute, Inc 2002) and SPSS version 17 for Windows statistical software packages were used for the analysis.

3. RESULTS

Pattern of variation among the genotypes were different for different traits. The largest variation was observed for fruit yield (t/ha), total fruit yield(kg), and fruit yield per plant(kg) with coefficients of variations of 75.47%,75.45% and 69.02% respectively. Generally, all the traits show moderate to high variability. Coefficient of variation (CV %) ranged from 13.52 to 75.47% for the various traits. The coefficient of variation was highest for fruit yield (t/ha), while the lowest level was showed by leaf length. Based on the agro-morphological characters, the papaya genotypes collected showed variation in most of the characters especially the fruit yield traits.

Table 1. Basic statistics of the agro-morphological characters of 10 genotypes of pawpaw

Characters	Mean	Mean Min. – Max.	Standard deviation	CV
Stem height (cm)	170.42	138.2 – 209.0	23.33	13.69
Stem girth (cm)	15.66	12.2 - 18.8	2.42	15.47
Leaf length (cm)	41.10	34.6 - 52.8	5.56	13.52
Leaf width (cm)	57.75	51.0 - 63.6	7.83	13.56
Petiole length (cm)	61.79	41.26 – 84.00	13.34	21.60
Fruit length (cm)	24.14	18.2 - 31.9	4.95	20.52
Fruit diameter(cm)	13.44	10.8 - 16.7	2.14	15.95
Fruit flesh thickness (cm)	3.03	2.3 - 3.5	0.42	14.00
No. of fruits per plant	14.60	8.0 - 26.0	4.88	33.43
Number of fruits Harvested	9.90	3.0 - 21.0	5.09	51.38
Total fruit yield (kg)	22.00	5.9 - 60.5	16.60	75.45
Fruit yield per plant (kg)	11.52	3.0 - 30.3	7.95	69.02
Fruit Yield (t/ha)	30.52	8.2 - 84.0	23.03	75.47

3.1 Principal Components Analysis (PCA)

The results showed that three principal components with eigen values more than one explained 93% of total variability (Table 2). The first ~~principal~~ component (PC1) as mostly fruit characters that explained 70.4% of total variability. Among the property vectors of PC1, leaf length, leaf width, fruit length , fruit per plant, number of fruit harvested~~-~~, total fruit yield, fruit yield per plant and fruit yield per hectare have higher values. The second principal component (PC2) is plant vegetative characters which explain 14.74% of total variability. Among the property vectors of PC2, stem height, stem girth, leaf length, leaf width and petiole length have higher values. The third principal component (PC3) is the remaining fruit characters that explain about 7.83% of total variability. All the agro-morphological characters in each component were positively correlated.

3.2 Cluster analysis

The dendrogram of the hierarchical cluster analysis (HCA) separated the 10 genotypes into different clusters with Squared Euclidean distance dissimilarities ranging between 3.09 to 108.31(Table not shown). Phenogram based on squared Euclidian distance coefficients using 13 traits placed the 10 genotypes into three main clusters (Figure 1). First cluster consisted of a total of four genotypes (40%) namely CP001, 011, 006 and 005. The second cluster consisted of five clusters while the remaining genotype (CP012) was in cluster 3. The Euclidean genetic distance between Cp006 and Cp012 was the highest (108.3) while the 92lowest Euclidean distance was between Cp001 and Cp011 (3.09). Except for stem height ,petiole lenght and fruit lenght, the third cluster (CP012) had highest values for the remaining ten characters (77% of the studied traits)

Table 2: Eigen values, variance, cumulative variance and component scores of the first three principal components for 13 quantitative traits in 10 papaya lines.

Traits	PC1	PC2	PC3
Stem height	-.081	.902	.182
Stem girth (cm)	.347	.681	.576
Leaf length (cm)	.575	.611	.447
Leaf width (cm)	.565	.569	.553
Petiole length (cm)	.392	.854	.091
Fruit length (cm)	.686	-.054	.596
Fruit diameter(cm)	.389	.525	.737
Fruit flesh thickness (cm)	.169	.235	.924
No. of fruits per plant	.826	.443	.182
Number of fruits Harvested	.967	.145	.004
Total fruit yield (kg)	.913	.182	.352
Fruit yield per plant (kg)	.895	.221	.375
Fruit Yield (t/ha)	.913	.182	.352
Eigen values	9.16	1.92	1.06
Cumulative eigen values	9.16	11.08	12.14
Variance (%)	70.44	14.74	7.83
Cumulative variance (%)	70.44	85.18	93.01

HIERARCHICALCLUSTER ANALYSIS

Dendrogram using Average Linkage (Between Groups)

Rescaled Distance Cluster Combine

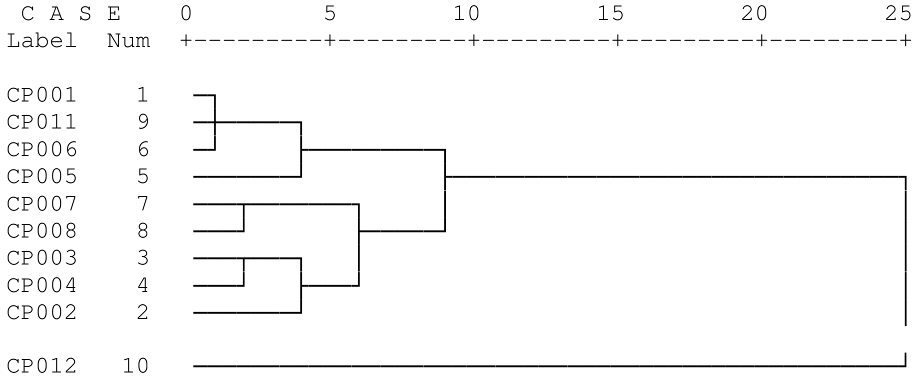

Figure 1: Dendrogram based on UPGMA analysis generated using the phenetic relationships among the 10 papaya lines.

3.3 Discriminant analysis

When discriminant function analysis was applied to group membership (the three clusters), only two of the agro-morphological characters (leaf width and fruit flesh thickness) was significant in distinguishing the cultivars. The leaf width was the most important character that discriminated the three groups from the cluster analysis followed by fruit flesh thickness. 100% of the original group's cases were correctly classified (Table 3).

Table 3. Discriminate function analysis: classification results

Classification Results[b,c]

			Predicted Group Membership			
		Groups	1	2	3	Total
Original	Count	1	4	0	0	4
		2	0	5	0	5
		3	0	0	1	1
	%	1	100.0	.0	.0	100.0
		2	.0	100.0	.0	100.0
		3	.0	.0	100.0	100.0
Cross-validated[a]	Count	1	4	0	0	4
		2	0	5	0	5
		3	0	1	0	1
	%	1	100.0	.0	.0	100.0
		2	.0	100.0	.0	100.0
		3	.0	100.0	.0	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 100.0% of original grouped cases correctly classified.

c. 90.0% of cross-validated grouped cases correctly classified.

4. DISCUSSION

All the 10 studied genotypes differed from one to another for all the characters. The range of values for most traits was high with the various genotypes having superior performance for each variable with CP012 particularly superior with respect to the studied traits and was the only genotype in its cluster group indicating that genotype of this cluster could be used as parent in future hybridization program for improved yield qualities. ~~Cluster analysis based on the agro-morphological characters resulted in three clusters~~. Crosses between individuals from different clusters may result in high heterosis. Even though, the genetic mechanisms that explain heterosis are not fully understood, it is well documented that crosses between unrelated and consequently genetically distant parents, show greater hybrid vigor than crosses between closely related parents [15] (Stuber 1994) since it is expected to produce new recombinants with desired traits. One of the important approaches to pawpaw breeding is hybridization and subsequent selection. Parents' choice is the first step in plant breeding program through hybridization. In order to benefit transgressive segregation, genetic distance between parents is necessary [16] (Joshi *et* al. 2004). The higher genetic distance between parents, the higher heterosis in progeny can be observed [17, 18](Joshi and Dhawan 1966; Anand and Murrty 1968). The genetic distance between CP006 and CP012 was the highest and thus crosses between these two parents are expected to produce new recombinants with desired traits. The principal component analysis indicated significant contributions in the component loadings of the 13 traits, which underpins their relevance in determining the variability among the 10 genotypes. The sign on the loadings indicates the direction of relationship between the components and the trait measured [19] (Biabani and Pakniyat 2008). Two traits with high weighting in the same component are expected to be highly correlated. This principle suggests that these traits could be probably influenced by similar gene(s) and may be used to identify variation among genotypes [19] (Biabani and Pakniyat 2008).

In spite of the reduction of the characters to only three principal components, it was possible to account for over 70% of the total variations among the date palm cultivars. Thus the capacity of PCA in data reduction without loss of information was confirmed [20] (Ross 1969). Component one loaded highly for fruit traits and accounted for over 70% of the total variation among the genotypes and therefore measured the importance of fruit characters in distinguishing the papaya genotypes. Leaf width as identified by discriminant analysis was important in distinguishing the pawpaw genotypes as it also loaded highly in all the three components in the principal component analysis.

In the present study, principal component analysis captured most of the variation within the genotypes in higher number of axes compared to discriminant analysis. Thus, a combination of PCA and discriminant analysis would be appropriate for describing the variation among papaya genotypes. [9] Odewale *et al*. (2012) also obtained similar result in coconut. Given the food and nutritional values of papaya, the morphological characterization of the papaya genotypes would serve as a good guide for the genetic development, conservation, collection and utilization of germplasm. Molecular studies would be useful to confirm the genetic diversity and characterize these genotypes for more detailed examination. This may help to emphasize the availability of these genetic resources for future breeding programmes.

5. CONCLUSION

The most divergent genotypes were CP006 and CP012 while the most similar genotypes were CP001 and CP011. Both multivariate methods showed similar results: Three clusters of genotypes were identified. The leaf width was the most important character that discriminated the three groups from the cluster analysis followed by fruit flesh thickness. Grouping of genotypes by multivariate methods in the study is of practical value for the papaya breeders. Representative genotypes may be chosen from the particular groups for hybridization programs with other approved cultivars. This will aid in identification, selection and combining genotypes to obtain important traits in one line with a broad genetic base. However further study across location and years needs to be done in order to corroborate the results obtained in the present investigation.

REFERENCES

1. Ramos HCC, Pereira MG, Pinto FO, Ribeiro EH. Multivariate analysis to determine the genetic distance among backcross papaya (*Carica papaya* L.) progenies. Genet. and Molecular Res. 2012; 11(2), 1280 – 1295.
2. Ming R, Yu Q, Moore PH. Sex determination in papaya. Semin. Cell Dev. Biol. 2007; 18, 401-408.

3. Van Droogenbroeck B, Breyne P, Goetghebeur P, Romeijn-Peeters E. AFLP analysis of genetic relationships among papaya and its wild relatives (Caricaceae) from Ecuador. Theor. Appl. Genet. 2002; 105, 289-297.
4. Krishna KI, Paridhav M, Jagruti AP. Review on nutritional, medicinal and pharmacological properties of papaya (*Carica papaya* linn.). National Prod. Reliance 2008; 7 (4), 364 – 373.
5. FAO (Food and Agriculture Organization) for the United Nations. :Statistics online website. http://faostat.fao.org, 2002; 2004; 2007; 2008.
6. Cruz CD, Carneiro PCS, Modelos biométricos aplicados ao melhoramento genético. - Viçosa: Imprensa Universitária. 2003; Vol. 2, 585pp.
7. **Dias LAS** and **Kageyama PY**. Multivariate genetic divergence and hybrid performance of cacao (*Theobroma cacao* L.). Braz. J. Genet. 1997; 20, 63-70.
8. Hamza AM, Agho C, Ado SG, Ikuenobe CE, Ataga CD Odewale JO. Proximate Compositions Evaluation and Variability among Cultivars of Date Palm (*Phoenix dactylifera* L.) in Nigeria. International Journal of Plant & Soil Science 3(3): 248-259, 2014
9. Odewale JO, Agho C, Ataga CD, Aisueni NO, Ikuenobe CE, Okoye MN, Odiowaya G, Edokpayi AA, Ahanor MJ, Uwadiae EO. Pattern of genetic diversity and variability in germplasm resources of local and exotic coconut (*Cocos nucifera L.*) cultivars in Nigeria. Scholarly J. Agric. Sci. 2012; 2(8), 202-207.
10. Singh P. Studies on genetic variability and diversity of rice. Madras Agric. J. 1983; 70(7), 436-440.
11. Jatasra DS, Parada RS. Genetic divergence in wheat under different environmental conditions. Cereal Res. Comm. 1978; 6, 307-317.
12. Zahan MI, Bhuiyan MSR, Hossain MS. Genetic divergence in Oleiferous Brassica species. J. Sher-e-Bangla Agric. Univ. 2008; 2(1), 1-6.
13. Kaiser HF. The application of electronic computers to factor analysis. Educational and Psychological Measurement. 1960; 20, 141–151. http://dx.doi.org/10.1177/001316446002000116
14. SAS Institute. Inc. SAS users's guide. Statistics, version 9.00. SAS Institute, Inc., Cary, NC, 2002.
15. Stuber CW. Heterosis in plant Breeding. Plant Breed. Rev. 1994; 12:227-251.
16. Joshi BK, Mudwari A, Bhatta MR, Ferrara GO. Genetic diversity in Nepalese wheat cultivars based on agro-morphological traits and coefficients of parentage. Nep. Agric. Res. J. 2004; 5, 7-17.
17. Joshi AB, Dhawan NL. Genetic improvement of yield with special reference to self-fertilizing crops. Indian J. Genet. And Plant Breed. 1966; 26,101-113.
18. Anand IJ, Murrty BR. Genetic divergence and hybrid performance in linseed. Int. J Genet plant Breed. 1968; 28, 178-185.
19. Biabani AR, Pakniyat H. Evaluation of seed yield-related characters in sesame (*Sesamum indicum* L.) using factor and path analysis. Pakistan J. Bio. Sci. 2008; 11, 1157-1160.
20. Ross CJS. Statistical algorithms. Algosithorus AS13-AS15. Appl. Stat. 1969; 18,103-110.