African Journal of Agricultural Research Vol. 7(17), pp. 2674-2678, 5 May, 2012
Available online at http://www.academicjournals.org/AJAR
DOI: 10.5897/AJAR11.1299
ISSN 1991-637X ©2012 Academic Journals

Full Length Research Paper

Determination of associations between three morphological and two cytological traits of yams (*Dioscorea* spp.) using canonical correlation analysis

P. E. Norman[1]*, P. Tongoona[2] and P. E. Shanahan[3]

[1]Njala Agricultural Research Center (NARC), PMB 540, Freetown, Sierra Leone.
[2]Africa Center for Crop Improvement (ACCI), University of KwaZulu-Natal, Private Bag X01, Scottsville 3209, South Africa.
[3]Agricultural Plant Sciences, University of KwaZulu-Natal, South Africa.

Accepted 30 March, 2012

Agro-morphological traits of plants may directly or indirectly depend on cytological traits. Thus, the determination of associations between morphological traits (absence or presence of wings, number of stems per plant and wing colour of stem) and cytological traits (DNA content and ploidy level) of yams were investigated using canonical correlation analysis. This multivariate technique is used in wide fields of study to quantify the mathematical relationships between multiple sets of independent and dependent traits or properties. Canonical weights and loadings indicated that DNA content (pg) had the highest contribution to the variation of the morphological traits (presence of wings, number of stems per plant and wing colour) compared with ploidy level. It was found that cytological traits accounted for 0.09 to 0.17% of the variation in the selected morphological traits. The first and second canonical correlations exhibited 60.91 and 39.09% overlapping variance of the canonical variate sets respectively. The first and second canonical variates extracted 0.57 and 4.43% of the total variance in the cytological trait set. The study demonstrated the successful determination of complex inter-relationships between morphological and cytological traits.

Key words: Canonical correlation, morphological traits, cytological traits, yam.

INTRODUCTION

Canonical correlation analysis (CCA) is one of several multivariate analysis techniques used to determine the overall correlation between two sets of traits (X and Y). Canonical correlation is a generalization of multiple regression analysis with more than one trait in the independent and dependent trait sets. The basic principle of the technique is to determine how much variance in one set of traits is accounted for by the other set along one or more axes (Tabachnick and Fidell, 2001). In contrast to many other techniques, any of the two sets of traits is a potential candidate to be used as dependent or independent traits. Canonical correlation makes possible several combinations of two trait sets. The number of combinations depends on the number of traits in the smaller trait set (Tabachnick and Fidell, 2001; Keskin and Yasar, 2007). Canonical correlation analysis has been widely applied in various fields such as the plant sciences, biology, chemistry, social and management sciences. However, there is scant information available on the interrelationship between morphological and cytological traits of yams. The main aim of this study was to determine the level of association between morphological and cytological traits of yams using canonical correlation analysis. The hypothesis being tested was that correlation exists between the agro-morphological and cytological traits used in the two methods of classification.

MATERIALS AND METHODS

A total of five traits including three morphological (absence or

*Corresponding author. E-mail: penorman2008@yahoo.com.

Table 1. Descriptive statistics of the cytological and morphological traits.

Traits	Mean ± SE	Minimum	Maximum
DNA (pg)	1.858 ± 0.013	1.588	2.118
Ploidy	3.962 ± 0.091	2.000	6.000
APW	0.846 ± 0.051	0.000	1.000
NS	1.846 ± 0.108	1.000	5.000
WC	1.404 ± 0.117	0.000	3.000

DNA: Deoxyribonucleic acid content (pg); Ploidy level; APW: absence or presence of wings; NS: number of stems per plant; and WC: wing colour.

presence of wings, number of stems per plant and wing colour of stem) and two cytological (DNA content and ploidy level) traits were used. The morphological traits were considered as the dependent Y-trait set, whereas the cytological traits were taken as the independent X-trait set. To obtain the maximum correlations between two sets of traits, two linear combinations were designed as follows:

$$W_i = a_{i1}X_1 + a_{i2}X_2 + \text{--------} + a_{ip}X_p \quad (1)$$

$$V_i = b_{i1}Y_1 + b_{i2}Y_2 + \text{--------} + b_{iq}Y_q \quad (2)$$

The symbols W and V represents canonical variates; a and b are canonical coefficients of the X and Y trait sets; and p (two traits) and q (three traits) are the number of traits in the X and Y trait sets, respectively. The estimation of the vector coefficients, a and b, was done according to Tabachnick and Fidell (2001).

To generate the canonical correlation for both sets of traits, the following formulae were used:

$$\text{var}(W) = a'\text{Cov}(X)\,a \quad (3)$$

$$\text{var}(V) = b'\text{Cov}(Y)\,b \quad (4)$$

$$C_{wv} = \frac{b'\text{Cov}(Y)\,b}{\sqrt{(a'\text{Cov}(X)\,a)(b'\text{Cov}(Y)b)}} \quad (5)$$

Where var (W) represents variance of the canonical variate W; var (V) is the variance of the covariate V; C_{wv} is the canonical correlation between the X and Y trait sets; Cov (Y) and Cov (X) are the covariances of the traits in the X and Y trait sets, respectively (Keskin and Yasar, 2007).

The relationship of a set of canonical variate is maximized when the correlation (r-value) of the p and q is small. The first set of canonical variate (W_1 and V_1) gives the highest correlation and is considered the most important. The correlation between W_2 and V_2 is only maximized where the traits measured are uncorrelated to W_1 and V_1. Similarly, the correlation between W_3 and V_3 is maximized if traits are not correlated with W_1, V_1, W_2 and V_2 (Manly, 1994). The canonical correlation analysis procedure (cancorrelation procedure) in Genstat version 12.1 was used to generate the relationships between sets of traits (Payne et al., 2009). The squared canonical correlation (also known as canonical roots or eigen-values) represents the amount of variance in one canonical variate accounted for by the other canonical variate (Hair et al., 1998). The standardized coefficients are similar to the standardized regression coefficients in multiple regression, which gives an indication of the relative importance of the independent traits in determining the value of dependent traits. In order to determine the amount of variance in one set of traits that is accounted for by another set of traits, Sharma (1996) suggested the estimation of the redundancy measure (RM) for each canonical correlation. The equation for the RM is shown as follows:

$$RM_{vi/wi} = AV(Y/V_i) \times C_i^2 \quad (6)$$

$$AV(Y/V_i) = [\sum^q LY_{ij}^2/q] \quad (7)$$

Where AV (Y/V_i) = the averaged variance in Y traits that is accounted for by the canonical variate V_i. LY_{ij}^2 = the loading of the jth Y trait on the ith canonical variate V_i; q = the number of traits in canonical variates; C_i^2 = the shared variance between V_i and W_i; W_i and V_i are canonical variates of Y and X trait sets, respectively.

This estimate is necessary because a large canonical correlation does not always imply powerful relationship between two sets of traits. Canonical correlation maximizes the estimate of correlation between linear combinations of traits in the two sets, but does not maximize the amount of variance accounted for in one set of traits by the other set of traits. Thus, the variance in one set of traits accounted for by the other set is obtained through the RM (Akbas and Takma, 2005).

RESULTS

The descriptive statistics and Pearson correlation coefficient (r) for the six traits are presented in Tables 1 and 2, respectively. The Pearson correlation coefficients for the traits ranged between -0.3928 and 0.7798 and were statistically significant ($p<0.05$), except for the association between APW and Ploidy (r = -0.2715) and between WC and NS (r = 0.1755) (Table 2). Although, the correlations between morphological and cytological traits were generally weak, the statistical significance of the correlations except between APW and ploidy indicated that the morphological traits were influenced by the cytological traits. In this study, the X trait set comprised of two traits: p = 2; and the Y trait set comprised of three traits: q = 3. Thus, two pairs of canonical variates, W_1V_1 and W_2V_2 were formed based on the set with the smaller number. The canonical correlations between these variates are presented in Table 3. Generally, results of canonical correlation indicated significant relationship between the two variable sets (Chipr 5, 191.109 < 0.001) with the highest contribution from U_1V_1. The significantly large value indicates that at least one of the r canonical correlations is significant. The first canonical correlation (W_1V_1) was 0.4441, which represents 69.91%

Table 2. Pearson correlation coefficients between cytological and morphological traits.

	DNA	Ploidy	APW	NS
Ploidy	0.7798**			
APW	-0.3928**	-0.2715^{ns}		
NS	0.2882*	0.3468**	-0.4727**	
WC	-0.3578**	-0.2895*	0.7143**	0.1755^{ns}

*: p< 0.05, **: p< 0.01, ns: not significant. DNA: deoxyribonucleic acid content (pg); Ploidy level; APW: absence or presence of wings; NS: number of stems per plant; and WC: wing colour.

Table 3. Canonical correlations between canonical variates.

Canonical variates	Canonical correlations	Squared canonical correlation	% correlation	Cumulative % correlation	Likelihood ratio	Probability Pr > F
W_1V_1	0.4441	0.1972	69.91	69.91	0.79	0.001
W_2V_2	0.2850	0.0812	39.09	100.0	0.89	0.005

Table 4. Non-standardized coefficients of the respective traits of the canonical variates.

Traits	W_1	W_2	Traits	V_1	V_2
DNA	0.9605	-2.2161	APW	-0.0436	0.5411
Ploidy	0.0909	0.3209	NS	0.1046	0.1724
			WC	-0.0999	-0.0931

DNA: Deoxyribonucleic acid content (pg); Ploidy level; APW: absence or presence of wings; NS: number of stems per plant; and WC: wing colour.

($\frac{0.4441}{0.4441 + 0.2850} \times 100$) of overlapping variance of the first canonical variate. The second canonical correlation W_2V_2, which exhibited 0.2850, represents 39.9% ($\frac{0.2850}{0.4441 + 0.2850} \times 100$) overlapping variance of the second canonical variate set (Table 3). The coefficients of canonical variates from the original data are presented in Table 4. These coefficients of canonical equations are not unique since the DNA coefficient value is more than 1.0. Therefore, the coefficients were standardized to give canonical variates with zero mean and unit variance.

The standardized canonical coefficients for the X and Y trait sets are presented in Table 5. The magnitude of each canonical coefficient represents the relative contribution of each trait to its respective canonical variate. From Equations 1 and 2, the following canonical variates can be obtained from the standardized coefficients (Table 5):

W_1 = 0.0890 DNA + 0.0591 Ploidy

V_1 = -0.0159 APW + 0.0815 NS – 0.0845 WC

W_2 = -0.2052 DNA + 0.2157 Ploidy

V_2 = 0.1971 APW + 0.1344 NS – 0.0787 WC

From the equations aforementioned, W_1 estimates the additive effect between DNA amount and ploidy level; whereas V_1 estimates the contrast between number of stems per plant on one hand and the other traits (wing colour and absence or presence of wing). This indicates that the large variation in morphological traits (wing colour and absence or presence of wing) compared to number of stems per plant was possibly due to the additive influence between the cytological traits, like DNA amount and ploidy level. However, the second canonical variate, W_2 estimates a contrast between DNA amount and ploidy level; whereas V_2 measures the difference between wing colour and the other traits (number of stem per plant and absence or presence of wing). This indicates that the variation in morphological traits (absence or presence of wing and number of stem per plant) compared to wing colour was possibly due to the influence of the cytological traits (DNA amount and ploidy level). Of the three morphological traits used, number of stems produced per plant and wing colour were more stable (their signs did not change in both canonical variates) compared to absence or presence of wing (Table 5). Ploidy level was also more stable compared to DNA content. This implies that a group of genotypes with similar ploidy level may not necessarily contain the same DNA content and consequently, the number of loci will

Table 5. Standardized coefficients of the respective traits of the canonical variates.

Traits	W_1	W_2	Traits	V_1	V_2
DNA	0.0890	-0.2052	APW	-0.0159	0.1971
Ploidy	0.0591	0.2157	NS	0.0815	0.1344
			WC	-0.0845	-0.0787

DNA: Deoxyribonucleic acid content (pg); Ploidy level; APW: absence or presence of wings; NS: number of stems per plant; and WC: wing colour.

vary. The proportion of the total variance extracted from a set of traits by a canonical variate of that set is equal to the quotient of the sum of square of loadings and the number of traits in the set. Thus, the first canonical variates, W_1 in the X trait set; and V_1 in the Y trait set, were estimated as 0.0057 [$(0.0890^2 + 0.0591^2)/2$] and 0.0047 [$(-0.0159^2 + 0.0815^2 + (-0.0845^2)/3$] respectively. Therefore, the first canonical variate (W_1) extracted 0.57% in the X trait set and 0.47% in the Y trait set.

The second canonical variates (W_2) in the X trait set and (V_2) in the Y trait set were estimated as 0.0443 [$(-0.2052^2 + 0.2157^2)/2$] and 0.0210 [$(0.1971^2 + 0.1344^2 + (-0.0787^2)/3$], respectively. The redundancy index in a canonical variate is expressed as the percentage of variance it extracts from its own set of traits. Thus, the first canonical variate (V_1) extracted 0.09% (0.0047×0.4441^2) of the variance in the X trait set; whereas, the second variate (V_2) extracted 0.17% (0.0210×0.2850^2) of the variance in the X trait set. The results suggest that traits in the Y trait set (APW, NS and WC) are influenced by those in the X trait set (DNA and ploidy).

DISCUSSION

The first pair of canonical variates (W_1V_1) had the highest (0.4441) estimated canonical correlation compared to the second pair of canonical variates [W_2V_2 (0.2850)]. The correlation between the first pair of canonical variate indicates that morphological traits: absence or presence of wing, number of stems per plant and wing colour are associated with cytological traits: DNA and Ploidy level. The signs of the standardized coefficients reflect the effects of DNA and ploidy on absence or presence of wing, number of stems and wing colour. Wright et al. (2008) suggested that the total amount of DNA in the genome (genome size) roughly reflects an estimate of the number of genes within a genome. Thus, an understanding of allelic diversity within germplasm is relevant in association with observed phenotypic variation. The redundancy estimates for the first and second canonical correlation suggested that 0.09 and 0.17% of the variance in the Y trait set (APW, NS and WC) was accounted by the X trait set (DNA and ploidy). Although, the percentages were small, the variation in each of the morphological traits showed a significant ($p<0.05$) input by cytological traits. It is possible that some morphological traits were more influenced by cytological traits than others. A future challenge would be to investigate the specific traits that are influenced by specific genes.

Conclusion

The associations between morphological and cytological traits of yams were investigated using canonical correlation analysis. The phenotypic expression of morphological traits was apparently influenced by cytological traits. The first and second canonical correlations exhibited 69.91 and 39.09% overlapping variance of the canonical variate sets, respectively. The first canonical variate extracted 0.57% of the total variance in the X trait set and 0.47% in the Y trait set. This study demonstrated the complex inter-relationships between morphological and cytological traits. This could be relevant as a pre-breeding guide to ploidy manipulation; and also for future investigation of the effect of specific genes on the phenotypic expression of genotypes.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial aid by Forum for Agricultural Research in Africa (FARA) under the African Development Bank Project – 'Promotion of Science and Technology for Agricultural Development in Africa,' and FARA grant No. 2008/FARA/EDU/MSc Fell/CORAF/019.

REFERENCES

Akbas Y, Takma C (2005). Canonical correlation analysis for studying the relationship between egg production traits and body weight, egg weight and age at sexual maturity in layers. Czech J. Anim. Sci., 50: 163-168.

Hair JF, Anderson RE, Tatham RH, Williams CB (1998). Multivariate Data Analysis. 5th Edition. Prentice Hall, Inc. pp. 442-462.

Keskin S, Yasar F (2007). Use of canonical correlation analysis for determination of relationships among several traits in egg plants (*Solanum melongana* L.) under salt stress. Pakistan J. Bot., 39: 1547-1552.

Manly BFJ (1994). *Multivariate statistical methods*. A Primer. Second edition. Chapman & Hall, 2-6 Boundary Row, London SE1 8HN, UK. pp. 214.

Payne RW, Murray DA, Harding SA, Baird DB, Souter DM (2009). Genstat for Windows (12th Edition) Introduction VSN International, Hemel Hempstead.
Sharma S (1996). Applied Multivariate Techniques. John Willey and Sons, Inc., Canada. pp. 391-404.
Tabachnick B, Fidell LS (2001). *Using Multivariate Statistics*. A Pearson Education Company, Needham Heights, USA. pp. 966.
Wright SI, Ness RW, Foxe JP, Barette SCH (2008). Genomic consequences of out-crossing and selfing in plants. Inter. J. Plant Sci., 169: 105-118.