ANALYSIS OF VARIATION AMONG CORCHORUS OLITORIUS (L) GENOTYPES BY FACTOR ANALYSIS AND PRINCIPAL COMPONENT ANALYSIS (PCA)

Sunday Clement Olubunmi Makinde  
Department of Botany  
Lagos State University, Ojo, Lagos, Nigeria

Regland Michael Onyemeka  
Department of Botany  
Lagos State University, Ojo, Lagos, Nigeria

Rachael Chinweoke Ogbuoka  
Department of Botany  
Lagos State University, Ojo, Lagos, Nigeria

Opeyemi Shakirat Oyetunji  
Department of Botany  
Lagos State University, Ojo, Lagos, Nigeria

Ifeoma Sussan Ezenwata  
Department of Biological Sciences  
Chukwuemeka Odumegwu Ojukwu University, Uli, Anambra, Nigeria

Muinat Abidemi Asuni  
Department of Botany  
Lagos State University, Ojo, Lagos, Nigeria

Abstract— Multivariate statistical methods are utilized to estimate accurate genetic diversity in crop breeding programmes. This study aimed at investigating genetic divergence in fifteen genotypes of Jute (Corchorus olitorius) and determines the characters responsible for the variation using Factor and Principal component analysis. The fifteen genotypes were collected from different locations within southern Nigeria. The experiment was carried out at the Lagos State University Botanical Garden using a Randomized Block Design (RBD) with three replications. The collected data were subjected to Factor Analysis and Principal Component Analysis (PCA) to evaluate the patterns of variation in these accessions. The PCA accounted for over 92% of the total variation in the first five PCs while Factor Analysis accounted for over 86% of the variation in the first four factors. Contributions’ of number of leaves per plant, plant height, number of branches, stipule length, leaf length, petiole length and blade length as identified by the two analysis methods leads to the conclusion that these traits contributes more to the total variation observed in the fifteen genotypes of Corchorus and therefore can be used in discriminating among the genotypes. The configuration of the genotypes along the axes of PC1 and PC2 identified genotype NG/179 as high yielding genotypes in terms of number of leaves and plant height and therefore can be selected directly. The results, as captured by the complementarity effect of the principal component analysis and factor analysis, suggest the existence of genetic variability among the genotypes.

Keywords: Corchorus olitorius, Multivariate, variability, Principal Component Analysis, Factor Analysis.

I. INTRODUCTION

The genus Corchorus belonging to the family Malvaceae with a chromosome number 2n=14. It is distributed throughout the tropical and subtropical regions of the world (Kundu, 1951; Purseglove, 1968; Chang and Miau, 1989). Corchorus species have been reported to be extremely variable morphologically, especially in the vegetative parts like the leaves (Edmonds, 1990). Although 215 species, subspecies, varieties, and forms have been reported under the genus Corchorus, precise number of good species is approximately 100 (Saunders, 2006). Makinde et al. (2009) stated that about 30 species of Corchorus is found in Africa. In Nigeria, Corchorus olitorius is most frequently cultivated as vegetable. It is also called bush okra, Jews mallows or Jute mallow in English. Some Nigerian names include Ewedu in Yoruba, Ahuhara in Igbo and Malafiya in Hausa. The two most common types of Corchorus in Nigeria are Oniyaya, widely branched with broad, deeply serrated leaves (Corchorus incisifolius) and Amughadu, a plant growing even taller with large finely serrated leaves (Corchorus olitorius) that are oblong in shape. In Cameroun and other west Arica counties, there are numerous local types varying among others in height, stem,
colour, leaf and fruit shapes (Ogunkanmi et al. 2010). Leaves of Corchorus are consumed as leafy vegetables in various parts of the world especially in Asia, the Middle East, and part of Africa. Besides, adding a distinct flavour to food, jute leaves act as thickeners in soups, stews, and sauces. The seed is also used as flavouring agents, and herbal tea is also made from the dried leaves. The leaves are rich in protein, β-carotene, iron, calcium, vitamin B and vitamin C. The folic acid content is subsequently higher than that of other folacin-rich vegetables (Chen and Saad, 1987; Duke 1983). Different plant parts of Corchorus can be used in non-orthodox medicine and may be used directly as pharmaceuticals. They may also serve as templates for chemical synthesis of bioactive principles (Hazra and Saha 2004). Corchorus species contain important bioactive compounds such as cardiac glycol-sides, strophanthinid, β-sitosterol, trpenoid-corosin, flavone glycoside, urasolic acid, vitamin C, β-carotene, mucilage and others are potential candidates for developing plant based drugs (Chopra et al. 1986; Sen, 2002). Based on these economic uses of the plant, there is need for improved varieties which can be achieved through crop improvement programmes involving large sample sizes of breeding materials (Odiyi et al. 2014).

For any meaningful crop breeding program, an accurate estimate of genetic diversity within and between gene pools is pre-requisite and multivariate statistical methods are utilized. When diversity is described on a multivariate criterion, multiple measurements are analysed simultaneously on each individual genotype under investigation (Odiyi et al. 2014). The commonly used methods are principal component analysis (PCA), cluster analysis, factor analysis, canonical analysis, discriminant analysis; Mahalanobis squared distance (D² Statistics) and multidimensional scaling. Multivariate statistical methods is useful because it determine the plant character which causes the diversity or dissimilarity to arise and the relative contributions that the various characters make to the total variability in the germplasm (Odiyi et al. 2014). Multivariate statistical methods have been successfully used in classifying, summarizing and describing variation patterns in populations of crop genotypes (Rhodes, and Martins, 1972; Ariyo et al. 1987, Raji, 2003; Nassir and Ariyo, 2007; Makinde, and Ariyo, 2010; Fayeun, and Odiyi, 2012; Cooley and Lohnes, 1971; Odiyi et al. 2014). The use of more than one technique to investigate genetic diversity in crops by many researchers is common because of complementary effects of the techniques on the analysis and it allows comparison among the techniques to know which one captures most of the variation and provide clearer and informative display of the relative positions of the genotypes. Factor Analysis and Principal Component Analysis are both ordination methods (Odiyi et al. 2014). Factor Analysis aside serving as means of identifying fundamental and meaningful dimensions of a multivariate set of data (Cooley and Lohnes, 1971), it assumes that a small number of observed construct are responsible for the correlation among a large number of observed variables (Bramel et al. 1984). Factor Analysis has been employed by various researchers in crop investigation to explain the observed relationship among numerous variables by determining the effect of each factor on the dependence structure (Rao and Paroda, 1982; Bramel et al. 1984; Accquaah et al. 1992; Ariyo, 1993; Nassir and Ariyo, 2007; Makinde, and Ariyo, 2010; Odiyi et al. 2014). Principal component analysis reduces data to clarify the relationship between two or more characters and partitions the total variance of the original characters into uncorrelated new variables (Wiley, 1981).

The study aimed at investigating the extent of genetic divergence in fifteen genotypes of Corchorus olitorius and determines characters responsible for the variation using Factor and Principal component analysis.

II. MATERIALS AND METHOD

The experiment was carried out at the Botanical Garden of the department of Botany, Lagos State University, Ojo, Lagos, Nigeria. **Plant materials:** Dry seeds of 15 accessions of Corchorus were collected from different centres in Nigeria; 9 of the accessions were collected from the National Centre for Genetic Resources and Biotechnology (NACGRAB) Ibadan. While seven genotypes were collected from farm lands in South East and South West, Nigeria. Table 1 contains the genotype coding and centres of collection. **Experimental design, procedure and management:** Attempt was made to break the seed dormancy by treating the seeds with hot water for ten seconds prior to sowing. Corchorus is a small seeded crop and in order to ensure that the seeds are evenly spread, the seeds of each genotype were mixed separately with fine river sand (1gm seed:10kg sand) and then drilled in rows on raised beds. The experimental design was Randomized Block Design (RBD) with three replicates. Each row was 15 meters long with spacing of 1m between two rows and within rows; the seedlings were thinned to a spacing of 25cm between plants. A pre-treatment of cured poultry was applied on all the plots before drilling of seeds. The plants were watered daily and monitored until they were fully established. **Data collection and analysis:** 10 quantitative data were collected on all the accessions eight-weeks after establishment. Three competitive plants from the two middle rows in each plot were harvested and observations were taken on the following quantitative characteristic; Plant height at flowering (cm), Number of leaves per plant at flowering, Leaf blade length (cm), Leaf petiole length (cm), Total leaf length (cm), Leaf width (cm), Stipule length (cm), Stem girth at maturity (cm), Numbers of branches per plant and Final plant height (cm).

The data obtained for each character were pulled together and the mean values obtained were subjected to analysis of variance, Principal component and Factor analysis using SAS.
Factor analysis was conducted for identifying important factors contributing to the total variation. The total variance and eigen values explained by factors are indicated in Table 2. Factors whose eigen values were greater than 1.0 were retained, which resulted in four factors. Characters with loadings greater than 0.4 in a factor were considered major. The analysis identified four factors out of which only the first four factors are considered important. These four factors accounted for more than 86% of the total variance. The first factor with eigen value of 4.18 explained 42.77% of the variance, second, third and fourth factors with eigen value of 2.20, 1.18 and 1.09 explained 21.97, 11.76 and 10.93% of the variance respectively. The scores of the major characters describing the first four factors are shown in Table 3. The first principal component axis which accounted for 42.77% of the total variation was significantly positively loaded with all the traits except petiole length, leaf width and stem girth. The second principal component axis was positively loaded for leaf related traits; petiole length (0.60), leaf length (0.41) and leaf width (0.55) and negatively loaded for stipule (-0.27) and number of branches (-0.26). The third principal component axis was positively loaded with number of branches (0.46), and stem girth (0.66). In the principal component axis four, blade length and leaf length were negatively loaded while stem girth and leaf width were positively loaded. Number of branches (0.74) and petiole length (0.38) were positively loaded in principal component axis five while for blade length (-0.35) and stem girth (-0.37) were negatively loaded.

The configuration of the 15 genotypes along the first three principal component axes is shown in figures 1, 2 and 3. The ordination of the genotypes on axes 1 and 2 (Figure 1) showed that NA/002 and ANB01 were most distant from all other genotypes. Also in Figure 2; NA/002 and ANB01 still differentiated and distant themselves from others. Likewise NG/179 distinguished and distant itself from others. Figure 3, showed another configuration of the accessions (axes 2 and 3); OY’01 and NG/40 also distinguished themselves from other genotypes; these two genotypes were described by leaf related characters (petiole length, leaf length and leaf width) that were associated with principal components 2 and 3.

IV. DISCUSSION

When dissimilarity between a pair of variety is defined on a multivariate criterion, it is useful to be able to determine the plant characters that cause the dissimilarity to arise and the relative contributions of the various characters that make to the total variability in the germplasm (Ariyo, 1993). From the result of the factor analysis and principal component analysis, it was clear that plant height at flowering, number of branches per plant, stem girth, blade length, leaf length, number of leaves per plant, petiole length and final plant height were important component of genetic variability among the genotypes. Importance of plant height at harvest and number of branches per plant has been reported in cowpea by (Ariyo, 1993 and Aremu et al. 2007) and in fluted pumpkin by (Odiyi et al. 2014). From the result of the factor analysis, it is possible that stem, branch and leaf traits shared some gene in common for their control. PC 1 makes it clear that plant height at flowering contributed to the leaf blade length, total leaf length, stipule length, number of branches at maturity as well as the number of leaves per plant at flowering.

Principal Component Analysis (PCA)
The PCA revealed that the first five factors of the ten principal components accounted for 92.67% of the variation. The scores of the major characters that described the first three principal component axes are presented in Table 3. The first principal component axis was positively loaded for leaf related traits; petiole length (0.60), leaf length (0.41) and leaf width (0.55) and negatively loaded for stipule (-0.27) and number of branches (-0.26). The second principal component axis was positively loaded for leaf related traits; petiole length (0.60), leaf length (0.41) and leaf width (0.55) and negatively loaded for stipule (-0.27) and number of branches (-0.26). The third principal component axis was positively loaded with number of branches (0.46), and stem girth (0.66). In the principal component axis four, blade length and leaf length were negatively loaded while stem girth and leaf width were positively loaded. Number of branches (0.74) and petiole length (0.38) were positively loaded in principal component axis five while for blade length (-0.35) and stem girth (-0.37) were negatively loaded.

The configuration of the 15 genotypes along the first three principal component axes is shown in figures 1, 2 and 3. The ordination of the genotypes on axes 1 and 2 (Figure 1) showed that NA/002 and ANB01 were most distant from all other genotypes. Also in Figure 2; NA/002 and ANB01 still differentiated and distant themselves from others. Likewise NG/179 distinguished and distant itself from others. Figure 3, showed another configuration of the accessions (axes 2 and 3); OY’01 and NG/40 also distinguished themselves from other genotypes; these two genotypes were described by leaf related characters (petiole length, leaf length and leaf width) that were associated with principal components 2 and 3.

IV. DISCUSSION

When dissimilarity between a pair of variety is defined on a multivariate criterion, it is useful to be able to determine the plant characters that cause the dissimilarity to arise and the relative contributions of the various characters that make to the total variability in the germplasm (Ariyo, 1993). From the result of the factor analysis and principal component analysis, it was clear that plant height at flowering, number of branches per plant, stem girth, blade length, leaf length, number of leaves per plant, petiole length and final plant height were important component of genetic variability among the genotypes. Importance of plant height at harvest and number of branches per plant has been reported in cowpea by (Ariyo, 1993 and Aremu et al. 2007) and in fluted pumpkin by (Odiyi et al. 2014). From the result of the factor analysis, it is possible that stem, branch and leaf traits shared some gene in common for their control. PC 1 makes it clear that plant height at flowering contributed to the leaf blade length, total leaf length, stipule length, number of branches at maturity as well as the number of leaves per plant at flowering.

Principal Component Analysis (PCA)
The PCA revealed that the first five factors of the ten principal components accounted for 92.67% of the variation. The scores of the major characters that described the first three principal component axes are presented in Table 3. The first principal component axis which accounted for 42.77% of the total variation was significantly positively loaded with all the traits except petiole length, leaf width and stem girth. The second principal component axis was positively loaded for leaf related traits; petiole length (0.60), leaf length (0.41) and leaf width (0.55) and negatively loaded for stipule (-0.27) and number of branches (-0.26). The third principal component axis was positively loaded with number of branches (0.46), and stem girth (0.66). In the principal component axis four, blade length and leaf length were negatively loaded while stem girth and leaf width were positively loaded. Number of branches (0.74) and petiole length (0.38) were positively loaded in principal component axis five while for blade length (-0.35) and stem girth (-0.37) were negatively loaded.

The configuration of the 15 genotypes along the first three principal component axes is shown in figures 1, 2 and 3. The ordination of the genotypes on axes 1 and 2 (Figure 1) showed that NA/002 and ANB01 were most distant from all other genotypes. Also in Figure 2; NA/002 and ANB01 still differentiated and distant themselves from others. Likewise NG/179 distinguished and distant itself from others. Figure 3, showed another configuration of the accessions (axes 2 and 3); OY’01 and NG/40 also distinguished themselves from other genotypes; these two genotypes were described by leaf related characters (petiole length, leaf length and leaf width) that were associated with principal components 2 and 3.
Table 2: Eigen values, factor scores, communality, percent and cumulative variance of 10 studied characters of *Corchorus* from the factor analysis.

<table>
<thead>
<tr>
<th>Characters</th>
<th>Factor1</th>
<th>Factor2</th>
<th>Factor3</th>
<th>Factor4</th>
<th>Communality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant height at flowering (cm)</td>
<td>0.94</td>
<td>-0.01</td>
<td>-0.20</td>
<td>0.25</td>
<td>0.99</td>
</tr>
<tr>
<td>Blade length (cm)</td>
<td>0.77</td>
<td>0.28</td>
<td>0.25</td>
<td>-0.36</td>
<td>0.86</td>
</tr>
<tr>
<td>Petiole length (cm)</td>
<td>0.01</td>
<td>0.86</td>
<td>0.18</td>
<td>-0.07</td>
<td>0.77</td>
</tr>
<tr>
<td>Total Leaf length (cm)</td>
<td>0.65</td>
<td>0.61</td>
<td>0.29</td>
<td>-0.33</td>
<td>0.98</td>
</tr>
<tr>
<td>Leaf blade width (cm)</td>
<td>-0.00</td>
<td>0.81</td>
<td>-0.25</td>
<td>0.39</td>
<td>0.87</td>
</tr>
<tr>
<td>Stipule length (cm)</td>
<td>0.64</td>
<td>-0.40</td>
<td>-0.13</td>
<td>-0.44</td>
<td>0.77</td>
</tr>
<tr>
<td>Number of branches/ plant at maturity</td>
<td>0.47</td>
<td>-0.38</td>
<td>0.50</td>
<td>0.01</td>
<td>0.62</td>
</tr>
<tr>
<td>Stem girth (cm)</td>
<td>0.12</td>
<td>-0.14</td>
<td>0.72</td>
<td>0.56</td>
<td>0.87</td>
</tr>
<tr>
<td>Final plant height (cm)</td>
<td>0.93</td>
<td>-0.03</td>
<td>-0.24</td>
<td>0.25</td>
<td>0.99</td>
</tr>
<tr>
<td>Number of leaves/plant at flowering</td>
<td>0.87</td>
<td>-0.16</td>
<td>-0.25</td>
<td>0.27</td>
<td>0.92</td>
</tr>
<tr>
<td>Eigen value</td>
<td>4.18</td>
<td>2.20</td>
<td>1.18</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>Percentage variance</td>
<td>42.77</td>
<td>21.97</td>
<td>11.76</td>
<td>10.93</td>
<td></td>
</tr>
<tr>
<td>Cumulative variance</td>
<td>41.77</td>
<td>63.74</td>
<td>75.50</td>
<td>86.43</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Eigen value, percentage variance accounted for and cumulative percentage for agronomic characters of the first five principal components.

<table>
<thead>
<tr>
<th>Characters</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant height before flowering (cm)</td>
<td>0.46</td>
<td>-0.00</td>
<td>-0.18</td>
<td>0.24</td>
<td>0.03</td>
</tr>
<tr>
<td>Blade length (cm)</td>
<td>0.38</td>
<td>0.17</td>
<td>0.23</td>
<td>-0.34</td>
<td>-0.35</td>
</tr>
<tr>
<td>Petiole length (cm)</td>
<td>0.01</td>
<td>0.60</td>
<td>0.16</td>
<td>-0.06</td>
<td>0.38</td>
</tr>
<tr>
<td>Leaf length (cm)</td>
<td>0.32</td>
<td>0.41</td>
<td>0.26</td>
<td>-0.31</td>
<td>-0.13</td>
</tr>
<tr>
<td>Leaf width (cm)</td>
<td>-0.01</td>
<td>0.55</td>
<td>-0.23</td>
<td>0.37</td>
<td>0.12</td>
</tr>
<tr>
<td>Stipule length (cm)</td>
<td>0.31</td>
<td>-0.27</td>
<td>-0.12</td>
<td>-0.42</td>
<td>0.01</td>
</tr>
<tr>
<td>Number of branches</td>
<td>0.23</td>
<td>-0.26</td>
<td>0.46</td>
<td>0.01</td>
<td>0.74</td>
</tr>
<tr>
<td>Stem girth (cm)</td>
<td>0.06</td>
<td>-0.09</td>
<td>0.66</td>
<td>0.54</td>
<td>-0.37</td>
</tr>
<tr>
<td>Final plant height (cm)</td>
<td>0.46</td>
<td>-0.02</td>
<td>-0.22</td>
<td>0.24</td>
<td>-0.08</td>
</tr>
<tr>
<td>Number of leaves</td>
<td>0.43</td>
<td>-0.11</td>
<td>-0.23</td>
<td>0.26</td>
<td>0.10</td>
</tr>
<tr>
<td>Eigen value</td>
<td>4.18</td>
<td>2.20</td>
<td>1.18</td>
<td>1.09</td>
<td>0.62</td>
</tr>
<tr>
<td>Proportion of variation (%)</td>
<td>42.77</td>
<td>21.97</td>
<td>11.76</td>
<td>10.93</td>
<td>6.24</td>
</tr>
<tr>
<td>Cumulative variance</td>
<td>41.77</td>
<td>63.74</td>
<td>75.50</td>
<td>86.43</td>
<td>92.67</td>
</tr>
</tbody>
</table>

Fig. 1: Configuration of the genotypes under axes 1 and 2.

Fig. 2: Configuration of fifteen genotypes under axes 1 and 3
Fig. 3: Configuration of fifteen genotypes under axes 2 and 3.

In PC 2, and PC3 the number of leaves per plant does not correlate with the plant height at flowering hence, it could be possible to select genotypes with more leaves without adversely affecting other economically important traits. A similar finding has also been reported in vegetable Amaranth (Shukla et al. 2009) and fluted pumpkin (Odiyi et al. 2014). Traits like stem girth and number of branches that were weak in PC1 became stronger in PC3, lost some strength in PC4, and while at PC5 number of branches became stronger while stem girth had little strength.

Factor analysis and principal component analysis identified some similar characters as the most important for classifying the variation among the Corchorus genotypes. These included; number of leaves, plant heights and number of branches. The similarity between the two techniques had been reported earlier in okra by (Ariyo, 1993), rice by (Nassir and Ariyo, 2007), groundnut by (Makinde, and Ariyo, 2010) and fluted pumpkin by (Odiyi et al. 2014). Although, the two techniques produced similar results, their underlying principles are substantially different from each other. While PCA does not rely on any statistical model and assumptions, factors analysis does. It is also important to note that factor analysis suffers from other drawbacks, such as absence of ‘error’ structure and the dependence upon scale used to measure the variables (Bartual et al. 1985). However, factor analysis captured all ten characters as important characters that discriminate the genotypes, compared to PCA which identified only plant height, number of leaves, number of branches, blade length and stipule length. The configuration of the genotypes along the axes of PC1 and PC2 shows that genotype NG/179 is high yielding genotypes in terms of number of leaves and plant height and therefore can be selected directly. Heterosis can be achieved when NG/179 is crossed with NG/0207 which is notable for higher number of branches.

The PCA accounted for over 92% of the total variation in the first five PCs while factor analysis accounted for over 86% of the variation in the first four factors. These findings correspond with the work of (Odiyi et al. 2014) on fluted pumpkin. Since no test of significance was performed for factor loadings, the decision was rather arbitrary as how many factors should be extracted from data set and what magnitude of loading coefficient a variable should possess to be considered important (Accquaah et al. 1992). Differences in results of multivariate techniques, with respect to characters which best summarized the within population variance, had earlier been reported by (Rao and Paroda, 1982; Nair et al. 1998; Nassir and Ariyo, 2007; Makinde, and Ariyo, 2010). As suggested by the workers, a combination of the identified characters will give a good description of the variability and hence discriminate among the genotypes.

V. CONCLUSION AND RECOMMENDATION

The results, as captured by the complementarity effect of the principal component analysis and factor analysis, suggest the existence of genetic variability among the genotypes. Factor analysis and principal component analysis jointly identified number of leaves, plant heights, number of branches, leaf blade length and petiole length as important characters that cause variation among the genotypes.

Heterosis can be achieved in subsequent breeding programmes when NG/179 is crossed with NG/0207 which is notable for higher number of branches.

VI. REFERENCE


