Cluster Analysis Using Quantitative, Qualitative and Molecular Traits
for the Study of the Genetic Diversity in Pineapple Genotypes

C. de Fatima Machado*, F.V.D. Souza, J.R.S. Cabral, C.A. da Silva Ledo,
A.P. de Matos and R. Ritzinger
Embrapa Mandioca e Fruticultura Tropical
CP 07, Cruz das Almas
Bahia, 44.380-000
Brazil

Keywords: Ananas spp., multivariate analysis, Gower algorithm, genetic distance, variability

Abstract
Cluster analysis using quantitative, qualitative and molecular variables is a useful tool in estimating genetic diversity between genotypes in germplasm collections. The objective of this study was to carry out a simultaneous analysis of quantitative, qualitative and molecular variables followed by clustering to study the genetic diversity between pineapple genotypes using the Gower algorithm. Eleven quantitative, five qualitative and forty-three molecular characteristics in ninety pineapple genotypes were evaluated. The cophenetic correlation coefficient of the joint analysis was higher when compared with the individual analysis coefficients. Ten groups were observed which indicates higher variability among genotypes evaluated. The simultaneous analysis of the quantitative, qualitative and molecular variables was efficient in the expression of the genetic diversity between pineapple genotypes when compared to the individual analysis for each type of variable.

INTRODUCTION
The characterization of germplasm can be accomplished through the use of phenotypic traits (morphologic and agronomic descriptors), passport data and molecular markers. The characterization of the accessions is important to determine the genetic variability, to identify duplicated accessions and to establish nuclear collections. The use of multivariate techniques is one of the factors that has urged the increase in the studies about genetic diversity among genotypes (Gonçalves et al., 2008, 2009; Mohammadi and Prasanna, 2003; Podany and Schmera, 2006; Sudré et al., 2007).

Cluster analysis of those traits are done individually, because the genetic distances are calculated according to the type of the trait. It has been reported by Cruz (2008) that the procedures to estimate dissimilarity measures can be based on quantitative traits (Euclidean distances or distances of Mahalanobis), binary traits (Jaccard index, Nei and Li index) and multicategorical traits (Cole-Rodgers et al., 1997). In that sense, several discrepancies are observed in relation to the clusters and to the inferences in relation to the quantification of the variability among accessions of a germplasm bank. A technique that allows the combined analysis of quantitative and qualitative data was proposed by Gower (1971).

The objective of this effort was to accomplish the characterization of pineapple accessions of the Active Germplasm Bank of Embrapa Cassava and Tropical Fruits through the combined analysis of quantitative, qualitative and molecular data.

MATERIALS AND METHODS
Ninety pineapple accessions of the Active Germplasm Bank of Embrapa Cassava and Tropical Fruits were evaluated. Each accession was evaluated with regard to 11 quantitative traits (fruit weight, crown weight, fruit length and diameter, crown length, axis diameter, soluble solids and acidity of the pulp, plant height, stalk length and

* cristina.machado@cnpmf.embrapa.br
diameter), 5 qualitative traits (leaf color, presence of spines along the borders of the leaves, fruit shape, colors of the fruit and the pulp) and 43 RAPD (Random Amplified Polymorphic - DNA) molecular markers.

A combined analysis of the quantitative, qualitative and molecular data was accomplished for determination of the genetic distance, based on the algorithm described by Gower (1971). The hierarchical clusters of the accessions were obtained by the UPGMA - Unweighted Pair-Group Method With Arithmetic Averages (Sneath and Sokal, 1973). The validation of the clusters was determined by the cophenetic correlation coefficient (Sokal and Rohlf, 1962) and its significance was calculated by the test of Mantel with 10,000 permutations (Mantel, 1967).

A statistical software was used (R Development Core Team, 2006) for the calculation of the algorithm of Gower (Gower, 1971). The cophenetic correlation coefficient was calculated using the Genes software (Cruz, 2008) and the dendrogram was generated based on the matrix distance using the software Statistica (Statsoft, 2005).

RESULTS AND DISCUSSION

The dendrogram of dissimilarity obtained through the algorithm of Gower (Gower, 1971) in the evaluation of 11 quantitative traits, 5 qualitative traits and 43 molecular markers is represented in Figure 1. The dissimilarity matrix mean calculated from the UPGMA method (0.15) provided the formation of 10 groups. On Table 1 the listed accessions belong to each one of the 10 formed groups.

The cophenetic correlation coefficient (CCC) was 0.92** (significant by the test of Mantel with 10,000 permutations). The high CCC value indicates high correlation between the original distance matrices and the cluster matrix; Mohammadi and Prasanna (2003), Podani and Schmera (2006) and Gonçalves et al. (2008) came to the same conclusion.

The analysis through the algorithm of Gower (Gower, 1971) was efficient in expressing the degree of genetic diversity among the pineapple accessions evaluated, demonstrating that the combined analysis of quantitative and qualitative data with the molecular markers provides a greater efficiency in the knowledge of the genetic diversity. According to Gonçalves et al. (2009), the type and number of the variables chosen can jeopardize the efficiency of the combined analysis, especially in the case of using a large number of data from molecular markers, in the quantification of the genetic diversity.

CONCLUSION

The pineapple accessions of the Active Germplasm Bank of Embrapa Cassava and Tropical Fruits show high genetic variability based on quantitative, qualitative and molecular traits.

Literature Cited


Tables

Table 1. Groups and accessions within groups formed in the cluster analysis obtained by the algorithm of Gower in the evaluation of 11 quantitative traits, 5 qualitative traits and 43 molecular markers.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Accessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>FRF-632</td>
</tr>
<tr>
<td>2</td>
<td>Guiana</td>
</tr>
<tr>
<td>3</td>
<td>Perolaera</td>
</tr>
<tr>
<td>4</td>
<td>RBR-1, SNG-3</td>
</tr>
<tr>
<td>5</td>
<td>SNG-2 (Quinari)</td>
</tr>
<tr>
<td>6</td>
<td>FRF-678, FRF-762, FRF-1399, Pao de Acucar</td>
</tr>
<tr>
<td>7</td>
<td>LBB-1384, LBB-1374</td>
</tr>
<tr>
<td>8</td>
<td>LBB-1385</td>
</tr>
<tr>
<td>9</td>
<td>Hibrido-3607, Natal Queen, Puerto Rico I-67, Smooth Cayenne, Smooth Cayenne 3, Smooth Cayenne 2</td>
</tr>
</tbody>
</table>
Fig. 1. Dendrogram of dissimilarity of 90 pineapple accessions obtained through the algorithm of Gower in the evaluation of 11 quantitative traits, 5 qualitative traits and 43 molecular markers.