

COMPARISON BETWEEN CLUSTER METHODS USING SIMULTANEOUS ANALYSIS OF QUALITATIVE AND QUANTITATIVE DATA IN BANANAS

Ledo, CAS¹, Roque, RL², Amorim, TB³, Gonçalves, ZS³, Ferreira, CF¹, Amorim, EP¹

carlos.ledo@embrapa.br

¹Embrapa Cassava and Fruits, Cruz das Almas, Bahia, Brazil

²Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brazil

³Universidade Federal do Recôncavo da Bahia, Cruz das Almas, Bahia, Brazil

Understanding the genetic variability in germplasm banks is essential to genetic breeding programs. Usually quantification of the variability is carried out using quantitative and qualitative data separately, which may not be enough to infer about the similarity between the genotypes evaluated. The objective of the present work is to compare methods of cluster analysis of banana genotypes considering quantitative and qualitative data, simultaneously and separately. Data from 19 banana genotypes using 23 guantitative and 67 molecular bands from SSR markers were used. The combined cluster analysis methods, INDOMIX, PCAMIX and PRINQUAL, were used. For the separate data, the cluster analysis considering the Mahalanobis distance was used for the quantitative data and the Nei and Li indice for the molecular data. The methods for the separate and combined data analysis were evaluated based on the proportion of erroneous classifications from the kmeans cluster analysis, considering the classification of the genotypes in five predefined groups. In general, the three methods used presented lower probability of erroneous classification when compared to the methods separately. The INDOMIX, PCAMIX and PRINQUAL methods may be used to quantify the variability present in situations when quantitative and qualitative data are considered simultaneously and showed to be more efficient when the cluster analysis was used considering quantitative and qualitative data, separately.