

**ANÁLISE DE AGRUPAMENTO UTILIZANDO VARIÁVEIS QUANTITATIVAS E QUALITATIVAS
PARA O ESTUDO DA DIVERSIDADE GENÉTICA EM GENÓTIPOS DE MANDIOCA SILVESTRE**

Carlos Alberto da Silva LEDO¹, Leônidas Francisco de Queiroz TAVARES FILHO¹,
Mayana Matos de OLIVEIRA², Thamyres Cardoso da SILVEIRA², Ariana Silva SANTOS²,
Alfredo Augusto Cunha ALVES¹, Leandro Simões Azeredo GONÇALVES³

RESUMO – A análise conjunta de variáveis qualitativas e quantitativas tem sido apontada como uma ferramenta útil na estimativa da diversidade genética entre os acessos de uma coleção de germoplasma. O objetivo deste trabalho foi promover a análise simultânea de variáveis qualitativas e quantitativas e posterior agrupamento para definição da diversidade genética entre genótipos de mandioca silvestre por meio do algoritmo de Gower. Foram avaliadas 4 características quantitativas e 10 qualitativas em 14 genótipos de mandioca silvestre. A análise simultânea de variáveis qualitativas e quantitativas foi eficiente em expressar o grau de diversidade genética entre os genótipos de mandioca silvestre quando comparadas com a análise individual para cada tipo de variável.

PALAVRAS-CHAVE: Análise multivariada, algoritmo de Gower, distância genética.

SUMMARY – CLUSTER ANALYSIS USING QUANTITATIVE AND QUALITATIVE VARIABLES FOR GENETIC DIVERSITY STUDIES IN WILD CASSAVA GENOTYPES. Cluster analysis using qualitative and quantitative variables has been demonstrated to be a useful tool in estimating genetic diversity between genotypes in germplasm collections. The objective of the present study was to carry out a simultaneous analysis of quantitative and qualitative variables followed by clustering for studying the genetic diversity between wild cassava genotypes using the Gower algorithm. Four quantitative and ten qualitative characteristics in 14 wild cassava genotypes were evaluated. The simultaneous analysis of the quantitative and qualitative variables was efficient in the expression of the genetic diversity between wild cassava genotypes when compared to the individual analysis for each type of variable.

KEYWORDS: Multivariate analysis, Gower algorithm, genetic distance.

¹ Embrapa Mandioca e Fruticultura Tropical, CP 07, Cruz das Almas, BA, CEP 44.380-000. E-mail: ledo@cnpmf.embrapa.br, leonidas76@gmail.com, aalves@cnpmf.embrapa.br

² Universidade Federal do Recôncavo da Bahia, Cruz das Almas, BA, CEP 44.380-000. E-mail: mayana.agr@hotmail.com, tchansilveira@hotmail.com, ana.silva0491@hotmail.com

³ Laboratório de Melhoramento Genético Vegetal, Centro de Ciências e Tecnologias Agropecuárias, Universidade Estadual do Norte Fluminense Darcy Ribeiro, Campos dos Goytacazes, RJ, CEP 28.013-600. E-mail: lsagrural@yahoo.com.br

INTRODUÇÃO

Para a caracterização de genótipos de mandioca silvestre faz-se necessário quantificar a variabilidade existente pela avaliação de variáveis quantitativas (agronômicas) e qualitativas (morfológicas e moleculares). O uso de técnicas multivariadas é um dos fatores que tem impulsionado o aumento nos estudos sobre diversidade genética entre genótipos. As análises de agrupamento dessas variáveis são feitas individualmente, visto que as distâncias genéticas são calculadas em função do tipo de variável utilizada. Cruz (2008) apresenta os procedimentos para estimar medidas de dissimilaridade com base em variáveis quantitativas (distâncias euclidianas ou de Mahalanobis), binárias (índice de Jaccard, Nei e Li etc.) e multicategóricas (Cole-Rodgers). Nesse sentido, observam-se várias discrepâncias em relação aos agrupamentos e às inferências em relação à quantificação da variabilidade entre acessos de um banco de germoplasma. Uma técnica que permite a análise simultânea de dados quantitativos e qualitativos foi proposta por Gower (1971). Este método permite que valores da matriz de distância fiquem compreendidos entre 0 e 1, sendo necessário a padronização das variáveis quantitativas e qualitativas. Este trabalho teve por objetivo avaliar a variabilidade genética entre genótipos de mandioca silvestre por meio da análise simultânea de variáveis quantitativas e qualitativas.

MATERIAL E MÉTODOS

Foram caracterizados 14 genótipos de mandioca silvestre utilizados como parentais no programa de melhoramento genético de mandioca da Embrapa Mandioca e Fruticultura Tropical. Avaliou-se 14 variáveis, sendo 4 variáveis quantitativas: comprimento do lóbulo central, largura do lóbulo central, relação comprimento/largura do lóbulo central e comprimento do pecíolo; e 10 variáveis qualitativas: cor da folha apical, cor do pecíolo, cor da folha desenvolvida, cor dos ramos terminais, cor da nervura, forma do lóbulo central, número de lóbulos, pubescência do broto apical, floração e sinuosidade. Os seguintes genótipos de mandioca silvestre foram avaliados: FLA 002-01, FLA 005-02, FLA 005-03, FLA 005-04, FLA 005-05, FLA 005-06, FLA 025V, FLA 029V, TOM A003-03, CMF 005, CMF 00X, CMF 001, ANO 007 e ANO 009.

Uma análise individual para as variáveis quantitativas e qualitativas foi realizada. Para as variáveis quantitativas calculou-se a distância euclidiana média e para as variáveis qualitativas utilizou-se a distância de Cole-Rodgers (Cole-Rodgers *et al.*, 1997). Para a análise simultânea das variáveis quantitativas e qualitativas utilizou-se o algoritmo de Gower (Gower, 1971). Os

agrupamentos hierárquicos das análises individuais e simultâneas a partir das matrizes de distância genética foram obtidos pelos métodos de UPGMA - *Unweighted Pair Group Method with Arithmetic Mean* (Sneath & Sokal 1973). A validação dos agrupamentos foi determinada pelo coeficiente de correlação cofenético (Sokal e Rohlf, 1962). A significância dos coeficientes de correlação cofenético foi calculada pelo teste de Mantel com 1000 permutações (Mantel, 1967).

Para a obtenção das matrizes de distância genética das análises individuais e cálculo dos coeficientes de correlação cofenético foi utilizado o programa Genes (Cruz, 2008). A matriz de distância genética utilizando o algoritmo de Gower foi obtido pelo programa R (R Development Core Team, 2006). O dendrograma foi obtido pelo programa Statistica 7.1 (Statsoft, 2005).

RESULTADOS E DISCUSSÃO

A análise simultânea das variáveis quantitativas e qualitativas pelo algoritmo de Gower apresentou o maior coeficiente de correlação cofenético entre a matriz de distância genética e a matriz de agrupamento, com valor de correlação de 0,95**. As análises individuais para as variáveis quantitativas e qualitativas apresentaram valores de 0,86** e 0,91**, respectivamente (Tabela 1). Os pontos de corte, definidos pela média da matriz de agrupamento, promoveu a formação de 3, 6 e 5 grupos, respectivamente para as análises baseadas nos dados quantitativos, qualitativos e conjunta.

Tabela 1. Coeficiente de correlação cofenético e número de grupos formados em função das matrizes de distâncias genéticas estimada para os dados das variáveis quantitativas, qualitativas e análise simultânea em 14 genótipos de mandioca silvestre.

Matriz de distância	Coeficiente de correlação cofenético	Número de grupos formados
Quantitativa	0,86**	3 ¹
Qualitativa	0,91**	6
Conjunta	0,95**	5

** Significativo a 1% de probabilidade pelo teste de Mantel com 1000 permutações. ¹ Baseado na média da matriz de agrupamento.

Observa-se que a matriz de distância conjunta apresentou valores de correlações altos e significativos com as matrizes de distância obtidas nas análises individuais, 0,77** e 0,96**, para as variáveis quantitativas e qualitativas, respectivamente (Tabela 2). A correlação entre as análises individuais apresentou valor médio e significativo de 0,60**.

Tabela 2. Correlação entre matrizes de distância genética estimada para os dados das variáveis quantitativas, qualitativas e análise simultânea em 14 genótipos de mandioca silvestre.

Matriz de distância	Quantitativa	Qualitativa	Conjunta
Quantitativa	1		
Qualitativa	0,60**	1	
Conjunta	0,77**	0,96**	1

** e * significativo a 1 e 5%, respectivamente, pelo teste de Mantel com 1000 permutações.

Portanto, a análise simultânea pelo algoritmo de Gower foi utilizada para representar graficamente a variabilidade dos genótipos de mandioca silvestre, pela sua maior eficiência em quantificar as diferenças entre eles. (Figura 1). Observa-se que o ponto de corte, média das distâncias da matriz de agrupamento, definiu a formação de 5 grupos: o grupo I formado pelos genótipos FLA 002-01, FLA 005-02, FLA 005-03, FLA 005-04, FLA 005-05, FLA 005-06, FLA 025V, FLA 029V; o grupo II formado pelos genótipos CMF 005 e CMF 00X; o grupo III formado pelo genótipo TOM A003-03; o grupo IV formado pelos genótipos ANO 007 e ANO 009; e o grupo V formado pelo genótipo CMF 001.

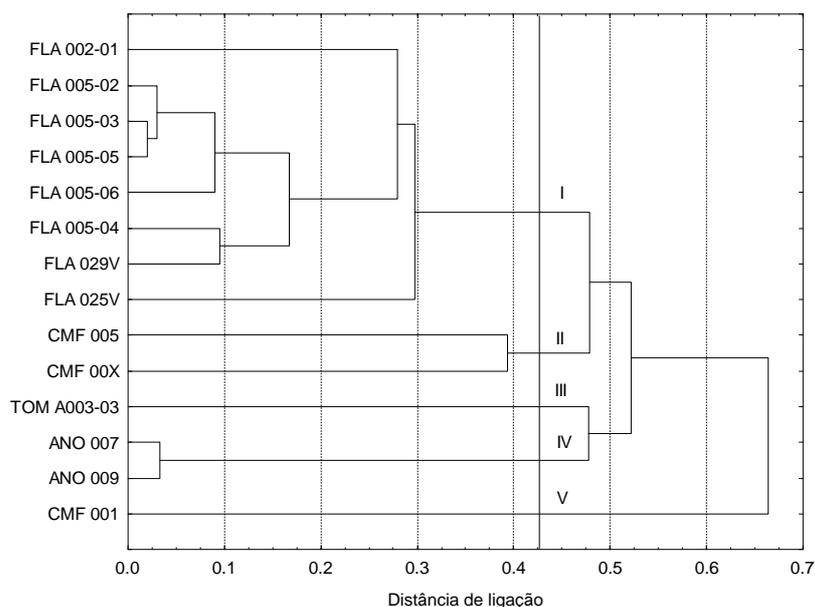


Figura 1. Dendrograma baseado na distância de Gower e método de agrupamento UPGMA na análise de 4 variáveis quantitativas e 10 variáveis qualitativas em 14 genótipos de mandioca silvestre.

A análise simultânea permite que, além de variáveis agronômicas e morfológicas, sejam utilizadas variáveis provenientes de marcadores moleculares. Segundo Gonçalves *et al.* (2009), a escolha das variáveis e o número de variáveis a serem usadas podem comprometer a eficiência da

análise simultânea, principalmente no caso de se utilizar um grande número de variáveis binárias, provenientes de marcadores moleculares, na quantificação da diversidade genética.

CONCLUSÕES

A análise pelo algoritmo de Gower foi eficiente em expressar o grau de diversidade genética entre os genótipos de mandioca silvestre avaliados, demonstrando que a análise simultânea proporciona uma maior eficiência no conhecimento da diversidade genética.

REFERÊNCIAS BIBLIOGRÁFICAS

- COLE-RODGERS, P.; SMITH, D. W.; BOSLAND, P. W. A novel statistical approach to analyze genetic resource evaluations using *Capsicum* as an example. *Crop Science*, v. 37, p. 1000-1002. 1997.
- CRUZ, C. D. *Programa genes* (versão Windows): aplicativo computacional em genética e estatística. Viçosa: UFV, 2008.
- GONÇALVES, L. S.; RODRIGUES, R.; AMARAL, A. T. JR.; KARASAWA, M; SUDRE, C. P. Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using a Ward-modified location model. *Genet. Mol. Res.*, v. 8, p. 364-374. 2009.
- GOWER, J. C. A general coefficient of similarity and some of its properties. *Biometrics*, Arlington, v. 27, n. 4, p. 857-874. 1971.
- MANTEL, N. The detection of disease clustering and generalized regression approach. *Cancer Research*, Birmingham, v.27, n.2, p.209-220, 1967.
- R DEVELOPMENT CORE TEAM. A language and environment for statistical computing. Vienna: *R Foundation for Statistical Computing*, 2006.
- SNEATH, P. H.; SOKAL, R. R. *Numerical taxonomy: The principles and practice of numerical classification*. San Francisco: W.H. Freeman, 1973. 573p.
- SOKAL, R. R. and ROHLF, F. J. The comparison of dendrograms by objective methods. *Taxon*, v.11 p. 33-40. 1962.
- STATSOFT, Inc. *Statistica for Windows (data analysis software system)*, version 7.1. Statsoft, Tulsa, Oklahoma (USA), 2005.