

## SELEÇÃO DE MARCADORES RAPD PARA O ESTUDO DA ESTRUTURA GENÉTICA DE POPULAÇÕES DE *Hancornia speciosa* Gomez

### SELECTION OF RAPD MARKERS TO STUDY GENETIC STRUCTURE OF *Hancornia speciosa* Gomez

Nara Fernandes MOURA<sup>1</sup>; Lázaro José CHAVES<sup>2</sup>; Roland VENCOSKY<sup>2,3</sup>; Maria Imaculada ZUCCHI<sup>4</sup>; José Baldin PINHEIRO<sup>2,3</sup>; Lizz Kezzy de MORAIS<sup>5</sup>; Mara Fernandes MOURA<sup>5</sup>

**RESUMO:** A mangabeira (*Hancornia speciosa* Gomez) é uma frutífera freqüentemente encontrada no cerrado brasileiro, bioma para o qual são escassas informações sobre a variabilidade genética de populações naturais. Sendo assim, o objetivo deste estudo foi a identificação de marcadores RAPD polimórficos que possibilitem a caracterização da estrutura genética de populações da espécie. A partir do DNA extraído de seis genótipos, foram analisados 45 “primers”, sendo 10 deles polimórficos, com os quais foram obtidos 60 locos polimórficos. Para se determinar o número mínimo de locos necessários para estudos moleculares em populações naturais de *H. speciosa*, utilizou-se o procedimento *bootstrap* para estimar o coeficiente de variação associado a número crescente de bandas. Tal parâmetro é estimado a partir do coeficiente de similaridade “*simple matching*” e suas respectivas dissimilaridades procedendo-se 1000 permutações. Os coeficientes de variação (y) obtidos foram relacionados com o número de locos (x) ajustando-se uma função potência do tipo  $y = ax^b$ . O número mínimo de locos foi determinado encontrando-se o valor de x correspondente ao ponto de curvatura máxima da equação ajustada. Verificou-se que, em pesquisas visando o número adequado de locos para alcançar uma dada precisão e tendo-se medidas de dissimilaridade ou similaridade, é recomendável utilizar, sempre a de maior média. Estes estudos permitirão investigar a divergência genética entre populações, contribuindo assim para pesquisas futuras sobre sua preservação e domesticação.

**UNITERMOS:** Mangabeira; Marcadores moleculares; Cerrado; Genética de populações.

### INTRODUÇÃO

A mangabeira (*Hancornia speciosa* Gomez) pertence à classe *Dicotyledoneae*, ordem *Gentianales* e à família *Apocynaceae*. De acordo com Lederman et al. (2000) a mangabeira é uma planta originária do Brasil e ocorre como frutífera nativa. A espécie *H. speciosa* é encontrada em várias regiões do Brasil, sendo nas áreas de cerrado do Centro-Oeste e nos tabuleiros costeiros e baixas litorâneas do Nordeste que a mangabeira apresenta em maior quantidade, e onde se processa o seu mais intenso extrativismo.

Devido à grande redução na área original dos ecossistemas em que a mangaba ocorre, principalmente

pelo desmatamento, especulação imobiliária e plantios como a cana-de-açúcar, o coqueiro e pastagens, a espécie é uma das frutíferas nativas mais ameaçadas de extinção no nordeste. Situação semelhante ocorre na região do Cerrado, que tem sido bastante devastada em função da expansão da fronteira agrícola para o cultivo intensivo de grãos e pastos (LEDERMAN et al. 2000).

Desta forma, é indispensável a definição de um programa de conservação de recursos genéticos que venha efetivamente dar condições para manutenção da espécie no local natural de ocorrência (*in situ*) ou fora do local (*ex situ*). Nesse particular, a caracterização genética é de fundamental importância, pois permitirá um melhor conhecimento dos recursos genéticos da região,

<sup>1</sup> Engenheira Agrônoma, Mestre, Escola de Agronomia, Universidade Federal de Goiás.

<sup>2</sup> Engenheiro Agrônomo, Doutor, Professor da Escola de Agronomia, Universidade Federal de Goiás.

<sup>3</sup> Agrônomo, Doutor, Professor da ESALQ, Universidade de São Paulo.

<sup>4</sup> Bióloga, Pós-Doutoranda, ESALQ, Universidade de São Paulo.

<sup>5</sup> Engenheira Agrônoma, Doutoranda, Escola de Agronomia, Universidade Federal de Goiás.

Received: 03/09/04

Accept: 28/0/05

e ainda, constituirá informação básica para programas de coleta, conservação de germoplasma, domesticação e melhoramento genético, para incorporação destas espécies nos sistemas produtivos regionais.

Assim, a conservação de recursos genéticos visa manter a variabilidade genética de espécies de interesse para o homem, para posterior utilização. Essa conservação exige um conhecimento detalhado da estrutura genética destas populações naturais, bem como de seus sistemas reprodutivos (LLERAS, 1992). Neste sentido, os marcadores moleculares têm sido freqüentemente utilizados em estudos da diversidade e estrutura genética populacional (FERREIRA; GRATTAPAGLIA, 1996). Os marcadores RAPD são muito úteis para mensurar similaridades ou distâncias genéticas e parâmetros afins, quantificando a variabilidade genética dentro e entre populações o que permite inferir sobre os possíveis meios de conservação da espécie.

Uma das principais etapas na implementação deste tipo de estudo reside na determinação do número de locos marcadores necessários para amostrar o genoma da espécie em questão. A utilização de um número de locos abaixo do necessário, poderá comprometer a precisão das inferências a serem feitas. Por outro lado, a utilização de um número elevado de locos marcadores poderá comprometer a amostragem do número de populações ou de indivíduos por população, por causa da limitação imposta pela capacidade laboratorial.

O método bootstrap tem sido recentemente empregado para responder a este tipo de questão. Este método tem uso bastante difundido para testar a significância das estatísticas calculadas através de dados moleculares, sendo baseado na reamostragem de dados reais para revelar algum padrão neles existentes. A principal virtude deste método é que produz tendências e erros-padrões de forma automática e fornece intervalos de confiança para as estatísticas calculadas (PEQUENO et al., 2003).

Diante da inexistência de um protocolo laboratorial para análise genética de populações de mangabeira utilizando RAPD, a pesquisa teve como objetivo estabelecer e divulgar o protocolo para extração de DNA, identificar marcadores RAPD e determinar o número de locos marcadores necessários para amostrar o genoma da espécie em questão, para futura caracterização genética de populações da espécie. Observa-se na literatura diversos estudos utilizando RAPD para análise da variabilidade genética em plantas e animais (TANKSLEY; BROWN, 2000; DEBERNHAM et al., 2000; HOFSTRA et al., 2000; JAGGI; WIRTH; BAUR, 2000; BOUZAT, 2001).

## MATERIAL E MÉTODOS

### Extração de DNA

Foram obtidas amostras foliares de seis genótipos localizados no arboreto da Escola de Agronomia da Universidade Federal de Goiás. A extração do DNA genômico total foi realizada a partir de folhas jovens de cada genótipo, com nitrogênio líquido, utilizando-se o método CTAB descrito por Murray e Thompson (1980) e Rogers e Bendich (1985), com modificações. Este método consiste em macerar, aproximadamente, 200 mg de tecido foliar fresco em tubos de 1,5 mL e incubação por 30 minutos a 65°C com 700 µl de tampão de extração (100 mM de tris-HCl, pH 8; 20 mM de EDTA; 1,4 M de NaCl; 2% de CTAB; 1% de PVP e 2% de mercaptoetanol) pré-aquecido. As proteínas foram eliminadas com solvente composto por clorofórmio, álcool isoamílico numa proporção de 24:1. Os ácidos nucleicos foram precipitados em isopropanol gelado e lavados com etanol a 70% e 90%, secos a temperatura ambiente e ressuspendidos em tampão TE (10 mM Tris-HCl pH 8; 0,1 mM EDTA).

### Quantificação de DNA

A concentração de DNA de cada amostra foi estimada por meio da comparação com uma amostra de concentração conhecida de DNA do fago λ íntegro, em gel de agarose 1,4%.

### Seleção de Primers

Foram testados os seguintes primers dos Kits A, B, C, E, F, G, H (Operon Technologies): A5, A6, A7, A8, A9, A10, B5, B10, B20, C1, C2, C3, C4, C8, C9, C11, C12, C13, C14, C15, C16, C17, C19, C20, E5, E8, F15, G1, G2, G3, G6, G12, G13, G14, H1, H2, H6, H9, H11, H12, H15, H16, H17, H18 e H19.

### Amplificação do DNA (PCR)

As amostras de DNA de cada genótipo foram utilizadas nas reações de amplificação. Cada reação consistiu de 25 µL contendo: 5 ng de DNA; 50 mM KCl, 10 mM Tris-HCl pH 8,3; 2,0 mM de MgCl<sub>2</sub>; 0,2 mM de dNTPs; 30 ng de "primers" e 2 unidades de Taq polimerase. As reações foram submetidas a 48 ciclos de amplificação após desnaturação inicial a 92°C por 4 minutos. Cada ciclo consistiu de 1 minuto a 92°C, 1 minuto e 30 segundos a 37°C e 1 minuto e 30 segundos a 72°C. Ao final de 48 ciclos, foi realizada uma extensão final de 5 minutos a 72°C.

### Eletroforese

Os fragmentos gerados por amplificação foram

separados por eletroforese em géis de agarose horizontal 1,4%, em uma concentração de 1% em tampão TBE (Trisborato 90 mM e EDTA 2 mM) para verificação de polimorfismo. Os géis foram corados com brometo de etídeo, visualizados sobre luz UV (ultra violeta) e fotografados em fotodocumentador EDA-Kodak. Os géis foram interpretados e cada indivíduo foi genotipado de acordo com a presença ou ausência de cada banda para todos os “primers” analisados, selecionando-se apenas as bandas mais nítidas e bem espaçadas. Foi gerada uma matriz de dados binários composta pela presença (1) e ausência (0) de bandas. Estes dados foram empregadas na análise estatística.

### Análise Estatística

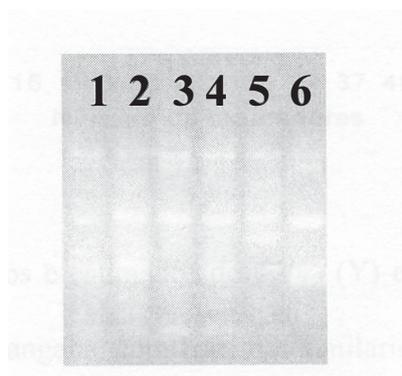
Com a finalidade de verificar se o número de locos polimórficos, detectados na pesquisa, foram suficientes para obter estimativas precisas das similaridades de concordância simples ( $S_{ij}$ , “Simple Matching”) e suas respectivas dissimilaridades ( $D_{ij} = 1 - S_{ij}$ ), aplicou-se o método de reamostragem (*bootstrap*) com número crescente de locos. Um critério para tomada de decisão, neste método, sobre a suficiência do número de locos (L) é o coeficiente de variação das estimativas (CV%) de  $S_{ij}$  e  $D_{ij}$  em cada nível de L. A razão de utilizar tanto as estimativas de  $S_{ij}$  como as de  $D_{ij}$  foi verificar como o aumento do número de locos afeta os valores de CV% destas duas medidas, bem como investigar se as conclusões sobre a suficiência do número de locos, para estimá-las, são coincidentes.

O processo de “bootstrap” foi realizado através do programa dBoot (COELHO, 2000). Em cada nível de L foram feitas até 1000 reamostragens. Com base nos valores de CV% (para  $S_{ij}$  e  $D_{ij}$ ) foram construídos gráficos de regressão para determinar um número mínimo de locos marcadores  $x$ . Os coeficientes de variação obtidos das duas medidas foram relacionados com o número de locos ajustando-se uma função potência do tipo  $y = ax^b$ . O número mínimo de locos foi determinado encontrando-se o valor de  $x$  correspondente ao ponto de curvatura máxima da função ajustada, utilizando-se a fórmula  $x_c = [a^2b^2(2b-1)/(b-2)]^{1/(2-2b)}$  (TRINDADE et al., 2001; PEQUENO et al., 2003).

A partir de uma matriz de similaridade baseada no índice *Simple Matching* foi construído um dendrograma pelo critério UPGMA.

### RESULTADOS E DISCUSSÃO

A partir do DNA extraído de seis (6) genótipos diferentes, foram analisados quarenta e cinco (45) “primers”, sendo selecionados como polimórficos os seguintes: A8, A9 (Figura 1), C4, C8, H2, H9, H10, H15, H18, H19 e estes revelaram sessenta (60) locos polimórficos. Conforme a Tabela 1, pode-se verificar a seqüência dos *primers* selecionados e o número de locos polimórficos gerados. O *primer* que produziu o maior número de locos polimórficos foi o H-15 e os que geraram o menor número de locos polimórficos foram o A-08 e o H-02.



**Figura 1.** Perfil de um gel de RAPD utilizando o *primer* Operon A-09 em 6 genótipos de mangabeira.

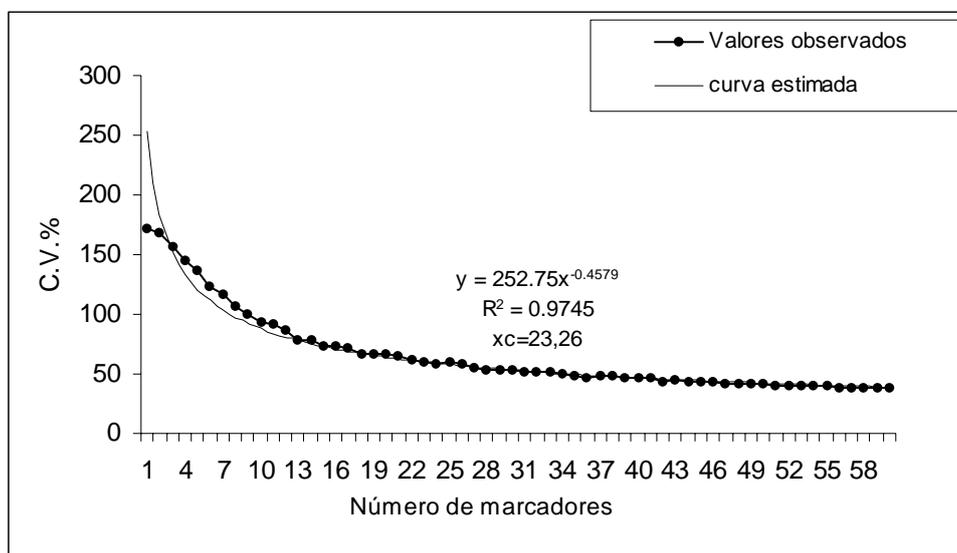
**Tabela 1.** Seqüência dos *primers* Operon selecionados para *H. speciosa* Gomez e número de locos polimórficos obtidos.

Primer	Seqüência 5'-3'	Número de locos polimórficos obtidos
A-08	TCGGACGTGA	3
A-09	GGGTAACGCC	6
C-04	CCGCATCTAG	7
C-08	TGGACCGGTG	5
H-02	GTGACGTAGG	3
H-09	TGTAGCTGGG	6
H-10	CCTACGTCAG	8
H-15	AATGGCGCAG	9
H-18	GAATCGGCCA	5
H-19	CTGACCAGCC	8

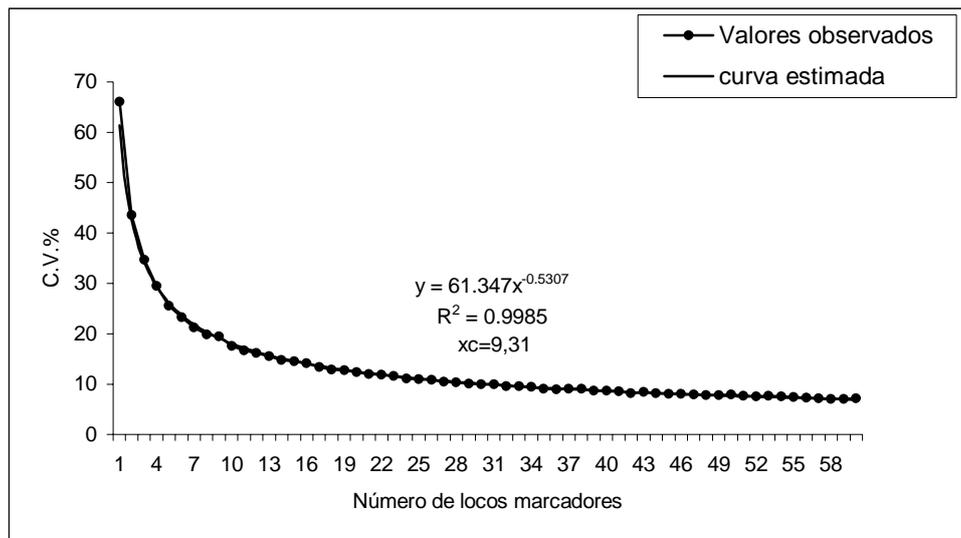
As curvas representativas das funções ajustadas, bem como os pontos referentes aos valores observados estão apresentados nas Figura 2 e 3. Ao realizar o procedimento de Bootstrap com as diferentes matrizes (de similaridades e dissimilaridades), os coeficientes de variação das estimativas de  $S_{ij}$  e  $D_{ij} = (1 - S_{ij})$  em cada nível de  $L$ , não foram coincidentes. Isto deve-se ao fato que o CV é estimado levando em consideração a média das similaridades e dissimilaridades obtidas pelas

reamostragens, assim se média das similaridades for muito pequena o CV conseqüentemente será elevado, e vice versa.

O coeficiente de determinação da função estimada, para a matriz de similaridades foi de 0,97 mostrando um bom ajuste dos valores observados à curva. Porém, o ponto de curvatura máxima da função ajustada foi  $x_c=23,26$ , correspondente a um CV de 59,60%.



**Figura 2.** Valores observados e estimados de CV% (Y) em função do número de locos marcadores (x) em mangaba, com base nas similaridades genéticas  $S_{ij}$  obtidas pelo método de “bootstrap” com 1000 reamostragens;  $x_c$  representa o ponto de curvatura máxima da função.



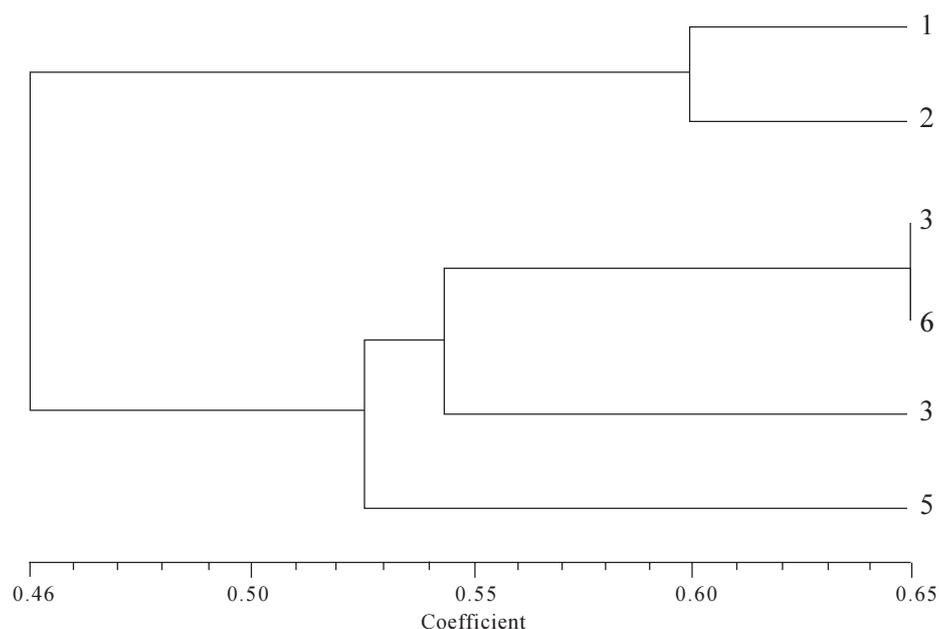
**Figura 3.** Valores observados e estimados de CV%(y) em função do número de locos marcadores (X) em mangaba, com base nas distâncias genéticas  $D_{ij}$  obtidas pelo método de *bootstrap* com 1000 reamostragens;  $x_c$  representa o ponto de curvatura máxima da função.

Os resultados obtidos indicam uma tendência clara de diminuição do coeficiente de variação na medida em que se aumenta o número de locos marcadores. Essa diminuição, porém tem o comportamento de uma curva assintótica ao eixo das ordenadas, significando que os decréscimos são proporcionalmente menores, na medida em que se aumenta o número de locos marcadores. Chega-se a um ponto em que o aumento do número de locos marcadores representa um aumento insignificante na precisão experimental, não justificando os acréscimos no esforço laboratorial (TRINDADE et al., 2001; PEQUENO et al., 2003). Este ponto significa o número mínimo de locos marcadores necessários para uma amostragem adequada do genoma, ou seja, 23,26 locos marcadores, no presente caso. A partir deste ponto a inclusão de novos locos marcadores vai depender da disponibilidade de recursos e do balanceamento entre a amostragem do genoma e a amostragem de plantas e populações, no campo, em função dos objetivos da pesquisa. Porém, estes resultados, demonstraram, que

quando as médias das similaridades obtidas por reamostragem são muito pequenas, o CV% é elevado e, conseqüentemente, não haverá decréscimo no CV% mesmo aumentando significativamente o número de locos marcadores.

No entanto, quando se compararam as curvas ajustadas para valores de similaridades e dissimilaridades, verificou-se que o ponto de curvatura máxima para valores de dissimilaridades foi bem menor  $x_c = 9,31$  locos, com um CV% correspondente de 18,77 e um coeficiente de determinação da curva de 99,5%. Este resultado indica que, quando se dispõe de indivíduos (populações) muito divergentes entre si, é recomendável utilizar matrizes de dissimilaridades para o procedimento de *Bootstrap*.

Pelo dendrograma (Figura 4) construído com o índice de similaridade *Simple Matching* pode se verificar que os indivíduos 3 e 6 são os mais similares (65%). Observa-se a formação de dois grupos, sendo um formado pelos indivíduos 1 e 2 e outro pelos indivíduos 3,4,5 e 6.



**Figura 4.** Padrão de similaridade genética obtido entre 6 genótipos de mangabeira, definido pelo critério de agrupamento UPGMA, utilizando o índice de similaridade de Jaccard.

## CONCLUSÕES

Existe polimorfismo de marcadores RAPD suficiente para o estudo da estrutura genética de populações de *Hancornia speciosa* Gomez. Além disso, com a presente pesquisa foi possível a implementação do protocolo laboratorial para extração de DNA e a identificação de marcadores RAPD para análise genética de populações de mangabeira utilizando RAPD.

Diante da diversidade genética detectada entre os genótipos avaliados, verificou-se que um número relativamente pequeno de locos RAPD foi suficiente para estimar dissimilaridades com suficiente precisão.

Em pesquisas visando determinar o número de locos necessário para estimar similaridades ou dissimilaridades, com uma precisão especificada e baseada no coeficiente de variação, é recomendável utilização da medida que tiver média maior.

---

**ABSTRACT:** *Hancornia speciosa* Gomez (mangabeira) is a fruit tree commonly found in the savanna regions (cerrado). Very little information is available about the genetic variability of the existing native populations. The objective of this work was to identify polymorphic RAPD markers for the characterization of the genetic structure of *H. speciosa* populations. The DNA was extracted from six individuals and analyzed using 45 set of primers. Ten of those primers were polymorphic and led to the identification of 60 polymorphic loci. In order to determine the minimum number of loci necessary for a study of this nature, the bootstrap procedure was used to obtain the coefficient of variation with increasing numbers of loci. These quantities were calculated for the simple matching coefficient of similarity and the corresponding dissimilarities through 1,000 permutations. Coefficients of variation (y) were related to the number of loci (x), according to the equation  $y = ax^b$ . The minimum number of loci was determined as the value of x corresponding to the maximum curvature of the adjusted equation. When researches are carried out to determine the required number of loci for attaining a given precision, and if similarities or dissimilarities are considered, the adequate procedure is to consider the quantity with the largest mean. These studies will permit broader investigations about the genetic divergence among populations and will contribute to further researches on the preservation and domestication of *H. speciosa*.

**UNITERMS:** Mangabeira; Molecular markers; Cerrado; Population genetic.

**REFERÊNCIAS BIBLIOGRÁFICAS**

BOUZAT, J. L. The population genetic structure of the greater Rhea (*Rhea americana*) in an agricultural landscape. **Biological Conservation**, Liverpool, v.99, n. 3, p.277-284, Jun., 2001.

COELHO, A. S. G. **dBoot**: avaliação dos erros associados a estimativas de distâncias/similaridades genéticas através do procedimento de bootstrap com número variável de marcadores. Goiânia: Laboratório de Genética Vegetal, Instituto de Ciências biológicas/UFG, 2000. CD.

DEBERNHAM, P.; BRZEZINSKI, M.; FOLTZ, K.; GAINES, S. Genetic structure of populations of the red sea urchin, *Strongylocentrotus franciscanus*. **Journal of Experimental Marine Biology and Ecology**, Groton, v.253, n.1, p.49-62, Oct., 2000.

FERREIRA, M.E.; GRATTAPAGLIA, D. Introdução ao uso de marcadores moleculares em análise genética. 2. ed. Brasília: Embrapa, 1996. 220 p.

HOFSTRA, D.E.; J. CLAYTON; J. D. GREEN; K. D. ADAM. RAPD profiling and isoenzyme analysis of New Zealand *Hydrilla verticillata*. **Aquatic Botany**, Gainesville, v.66, n.2, p.153-166. Feb., 2000.

JÄGGI, C.; T. WIRTH; B. BAUR. Genetic variability in subpopulations of the asp viper (*Vipera aspis*) in the Swiss Jura Mountains: implications for a conservation strategy. **Biological Conservation**, Liverpool, v.94, n.1, p.69-77, Jun., 2000.

LEDERMAN, I. E.; SILVA JÚNIOR, J. F. da; BEZERRA, J. E. F.; ESPÍNDOLA, A. C. Mangaba (*Hancornia speciosa* Gomez). In: LEDERMAN, I. E. (Ed.) **Série frutas nativas**. Jaboticabal: Funep, 2000. 35p.

LLERAS, E. Conservação de recursos genéticos florestais. **Revista do Instituto Florestal**, São Paulo, v. 4, p.1179-1184. 1992.

MURRAY, M. G. & W. F. THOMPSON. Rapid isolation of high molecular weight plant DNA. **Nucleic Acids Research**, Oxford, v. 8, n. 19, p. 4321-4325, Oct., 1980.

PEQUENO, S. A.; PINHEIRO, J. B.; ZUCCHI, M. I.; VENCOVSKY, R.; COELHO, A. S. G.; TRINDADE, M. G. Determinação do número de marcadores RAPD para estudos da diversidade genética em soja utilizando o método Bootstrap. **Bioscience Journal**, Uberlândia, v.19, n. 2, p. 45-48, May/Aug., 2003.

ROGERS, S. O.; BENDICH, A. J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plants tissues. **Plant Molecular Biology**, Dordrecht, v.5, n.2, p.69-76. Mar., 1985.

TANSLEY, S. A.; C. R. BROWN. RAPD variation in the rare and endangered *Leucadendron elimensen* (Proteaceae): implications for their conservation. **Biological Conservation**, Liverpool, v.95, n.1, p.39-48. Aug., 2000.

TRINDADE, M. G.; CHAVES, L. J.; COELHO, A. S. G.; ZUCCHI, M. I. Determinação do número de marcadores RAPD para estudos de estrutura genética em populações de *Eugênia dysenterica* utilizando o método bootstrap. In: CONGRESSO BRASILEIRO DE MELHORAMENTO DE PLANTAS, 1. 2001, Goiânia. **Anais do Congresso Brasileiro de Melhoramento de Plantas**. Goiânia-GO: SBMP, 2001. (CD).