

GRUPOS DE DIVERSIDADE GENÉTICA EM *Manihot esculenta* Crantz COM BASE NA ANÁLISE DISCRIMINANTE DE COMPONENTES PRINCIPAIS

Eder Jorge de Oliveira⁽¹⁾, Vanderlei da Silva Santos⁽¹⁾, Cláudia Fortes Ferreira⁽¹⁾, Gilmara Alvarenga Fachardo Oliveira⁽²⁾, Maiane Suzarte da Silva⁽²⁾, Carlos Ivan Aguiar-Vildoso⁽³⁾

⁽¹⁾Pesquisador da Embrapa Mandioca e Fruticultura, Rua da Embrapa s/n, CP007, Cruz das Almas (BA), E-mail: eder@cnpmf.embrapa.br, vssantos@cnpmf.embrapa.br, claudiaf@cnpmf.embrapa.br;

⁽²⁾Estudante de Biologia da Universidade Federal do Recôncavo da Bahia, Campus de Cruz das Almas (BA), E-mail: gfachardo@yahoo.com.br, maisuzarte@yahoo.com.br; ⁽³⁾Bolsista PNPd CNPq/Embrapa Mandioca e Fruticultura, E-mail: vildoso@hotmail.com

Introdução

Análise das relações genéticas existentes em diferentes espécies constitui-se em uma importante etapa para o desenvolvimento de programas de melhoramento, pois fornece informações sobre a diversidade genética e estratificação das populações de melhoramento. Avaliação precisa dos padrões de diversidade genética contribui para a geração de informações sobre: a) a variabilidade genética em cultivares; b) identificação de combinações parentais que permitam a geração de progênie segregantes com máxima variabilidade genética; e c) introgressão de genes desejáveis presentes nos diversos tipos de germoplasma disponível.

Recentemente métodos adequados de classificação e ordenamento da variabilidade genética tem assumido grande importância nos trabalhos de conservação e caracterização de germoplasma, e em programas de melhoramento genético, em função do aumento do número de genótipos e variáveis amostrados. Análises multivariadas, tais como análise de componentes principais (ACP) têm se mostrado eficientes em extrair informações dos mais diversos tipos de variáveis devido à sua capacidade de resumir a informação genética multivariada em algumas variáveis sintéticas.

Este tipo de análise possui vantagens sobre outras abordagens clássicas como o agrupamento Bayesiano implementado no software *Structure* (Pritchard et al., 2000), tais como: a) caráter exploratório que não exige adequações a estruturas genéticas como equilíbrio de Hardy-Weinberg ou a ausência de desequilíbrio de ligação; b) não serem computacionalmente intensivos, podendo ser utilizados para grandes conjuntos de dados; e c) permite abordar questões complexas como a identificação de adaptação, associando a variabilidade genética a condições ambientais específicas (Jombart et al., 2009). Entretanto, a ACP não permite inferências sobre a formação dos grupos de diversidade, exigindo uma definição a priori para estudar a estrutura da população, além disso, não é apropriado para uma imagem clara da

variação entre populações. Por outro lado, esta é a proposta da análise discriminante (DA), que define um modelo no qual se cria variáveis sintéticas com a variação genética particionada dentro e entre grupos, de forma maximizar a primeiro tipo de variação (Jombart et al., 2010).

O objetivo deste trabalho foi implementar a análise discriminante de componentes principais (ADCP), que associa as vantagens das duas metodologias multivariadas, para analisar e definir o número de grupos de diversidade genética no germoplasma de mandioca (*Manihot esculenta* Crantz.).

Material e Métodos

Dados de genotipagem

Para a definição dos grupos de diversidade em mandioca, foram analisados dados de genotipagem de 1280 acessos de germoplasma do Banco Ativo de Germoplasma de Mandioca (BAG-Mandioca) da Embrapa Mandioca e Fruticultura, avaliados com 402 SNPs.

Análise dos dados

A análise discriminante de componentes principais (ADCP) disponível no pacote *adegenet* para R, versão 2.11 (R desenvolvimento Core Team 2010) foi utilizada para definição dos agrupamentos, pois esta técnica não requer uma definição a priori de grupos genéticos (Jombart et al., 2010). Foram utilizados sucessivos agrupamento com o método *K-means* e o Critério de Informação Bayesiano (BIC) para definição do número de grupos, em que *K* com o menor valor BIC representa o número de grupos mais provável para o conjunto de dados em análise. Contudo, na presença de estruturação genética em cline (tipo particular de variação contínua) ou de forma hierárquica, os valores de BIC podem ser reduzidos após a identificação do verdadeiro valor de *K*. Assim, a redução dos valores BIC foi analisada visualmente para identificar o valor de *K* na qual os valores BIC reduziram apenas sutilmente (Jombart et al., 2010).

Foram testados valores de *K* de 1 a 50, com 10 corridas para cada *K*. Após definido o número de grupos, foram retidos os eixos da análise de componentes principais que explicam mais de 90% da variância total dos dados.

Resultados e Discussão

A ADCP foi utilizada para investigar o padrão de diversidade genética do germoplasma de mandioca da Embrapa Mandioca e Fruticultura. Foram mantidos na análise 200 componentes principais na etapa preliminar de transformação de dados, que no conjunto geral explicou mais de 90% da variação genética total (Figura 1). A definição do número de CPs a serem retidos na análise é um ponto de discussão sobre o poder de redução na dimensionalidade dos dados. De modo geral, diversos trabalhos indicam o uso de componentes que retenham mais de 80% da variância genética. No contexto da ADCP é preciso definir um ponto de equilíbrio entre o poder de discriminação dos

agrupamentos e a estabilidade de atribuições dos genótipos em cada grupo. Assim, a análise com 200 CPs garante alto poder estatístico para avaliar a estrutura genética do germoplasma de mandioca.

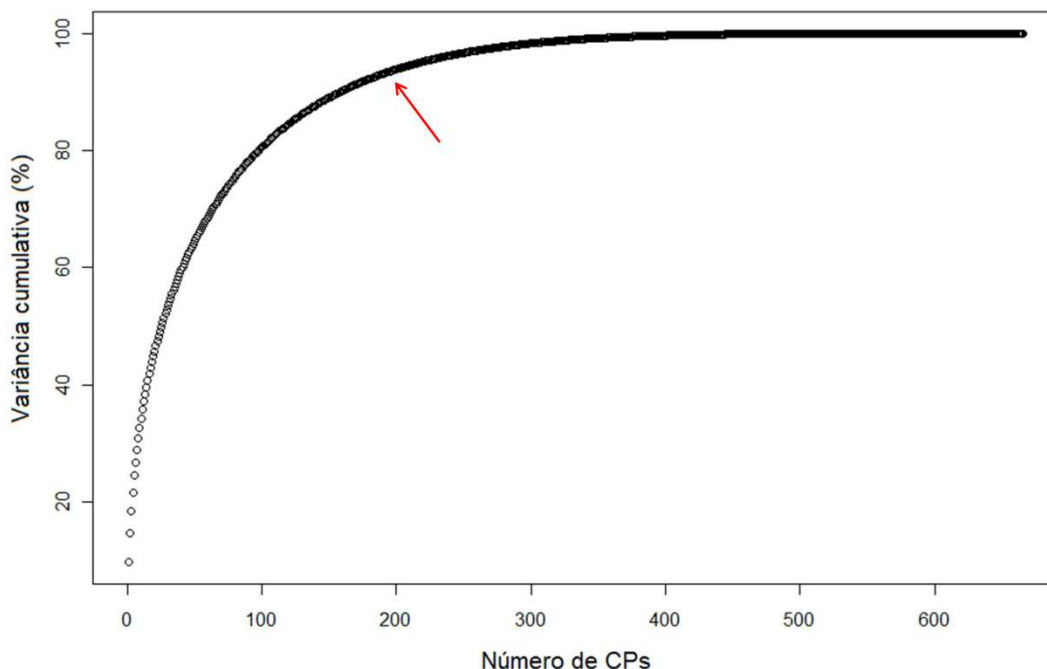


Figura 1. Distribuição cumulativa da variância para os componentes principais da análise de 1280 acessos de mandioca, genotipados com 402 marcadores do tipo SNPs.

A Figura 2 indica a presença de estrutura hierárquica no germoplasma de mandioca da Embrapa Mandioca e Fruticultura e ilustra o processo de seleção do número ideal de grupos. No agrupamento *K-means*, a queda inicial mais acentuada de acordo com o Critério de Informação Bayesiano (BIC) variou de $K = 30$ a 34 (Figura 2). Assim, foi adotado $K = 32$ para representar a variabilidade do germoplasma de mandioca. De acordo com Jombart et al. (2010), o tipo de estruturação da população influencia na precisão do método em relação à definição do número de agrupamentos, sendo que inferências em populações estruturadas no modelo de ilha são mais precisas do que em populações contínuas, como parece ser o caso do germoplasma de mandioca. Porém, mesmo quando o K real não é identificado, o número de agrupamentos inferido geralmente é muito próximo ao valor verdadeiro.

A estruturação do BAG-Mandioca em 32 grupos de diversidade abre possibilidades para exploração do efeito heterótico entre os agrupamentos formados. Neste caso, parentais com alta diversidade genética nos grupos, podem ser utilizados em cruzamentos dialélicos visando a implementação da seleção recorrente, ou mesmo a maximização do potencial de geração de populações superiores para testes clonais de competição.

A clara definição do número de agrupamentos no germoplasma de mandioca, com uso da estratégia de ADPC constitui-se em avanço metodológico, pois até o momento apenas métodos

hierárquicos ou apenas análise de componentes principais era utilizada para os estudos genéticos nesta cultura. Assim, a formação dos agrupamentos sempre foi feita de forma empírica, sem critérios definidos para a classificação e ordenamentos dos genótipos.

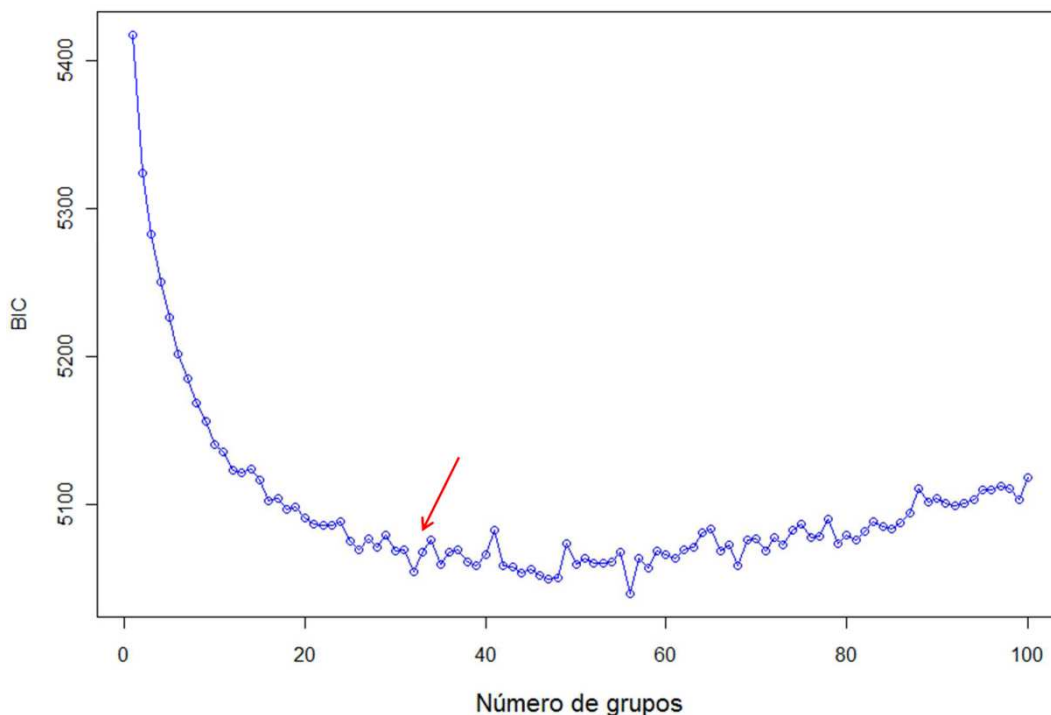


Figura 2. Distribuição das diferenças entre as heterozigosidade esperadas e observadas na análise de 1280 acessos de mandioca com 402 marcadores do tipo SNP.

Segundo Jombart et al. (2010), a ADCP pode ser útil para uma grande variedade de organismos, independentemente da sua ploidia e taxa de recombinação genética, pois a metodologia é independente de qualquer modelo de genética de populações, sendo portanto livre de suposições sobre equilíbrio de Hardy-Weinberg ou desequilíbrio de ligação. Mesmo em relação às abordagens de agrupamento bayesiana, a ADCP apresenta a vantagem de permitir a geração de uma representação gráfica do parentesco entre os agrupamentos formados.

Conclusões

A Análise Discriminante de Componentes Principais (ADCP) fornece um critério objetivo para a definição de grupos de diversidade genética no germoplasma de mandioca, contribuindo para avaliar a complexa estruturação da variabilidade desta espécie.

Agradecimentos

Ao CNPq e Fapesb pelo apoio financeiro.

Referências

JOMBART, T.; DEVILLARD, S.; BALLOUX, F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. **BMC Genetics**, v.11, 94, 2010.

JOMBART, T.; PONTIER, D.; DUFOUR, A.B. Genetic markers in the playground of multivariate analysis. **Heredity**, v.102, p.330–341, 2009.

PRITCHARD, J.K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v.155, p.945–959, 2000.