

Aplicação da PCAM e da análise de agrupamento em clones de cupuaçuzeiro

Vinicius Silva dos Santos^{1 3}

Sebastião Martins Filho¹

Rafael Moysés Alves^{2 4}

1 Introdução

O cupuaçuzeiro *Theobroma grandiflorum* (Willd. ex Spreng.) Schum. é uma fruteira nativa da região amazônica com destaque na agroindústria. Nos programas de melhoramento genético do cupuaçuzeiro, o estudo da diversidade genética tem sido de fundamental importância, pois visa selecionar, dentro da variabilidade dos genótipos, aqueles superiores, quanto à produção de frutos e resistência às doenças (ALVES et al., 2010). Como o cupuaçuzeiro é uma espécie perene com ciclo reprodutivo longo, torna-se necessária a utilização de métodos de seleção mais precisos. Nesta situação, o procedimento ótimo de seleção é o que envolve a estimação de componentes de variância pelo método da máxima verossimilhança restrita (REML) e a predição dos valores genotípicos pela melhor predição linear não-viesada (BLUP)(Resende, 2007a).

Para estudar a diversidade genética, faz-se necessário o uso de técnicas multivariadas. Dentre as mais empregadas estão a análise de componentes principais, as de variáveis canônicas e os métodos de agrupamento hierárquicos e de otimização. A análise de componentes principais (PCA) baseada na metodologia de modelos mistos padrão é denominada PCAM (PCA sob modelo misto)(Resende, 2007a). O objetivo deste trabalho foi estudar a divergência genética de 50 clones de cupuaçuzeiro, por meio da PCAM e da Análise de Agrupamento.

2 Material e Métodos

O experimento foi instalado em fevereiro de 1997, na Base Física da Embrapa Amazônia Oriental, em Tomé Açu, Pará. Foram avaliados cinquenta clones de cupuaçuzeiro em delineamento inteiramente ao acaso com dez repetições. Os dados foram obtidos com base na média de cinco frutos por planta nas safras de 2004 a 2011 e os seguintes descritores de frutos foram avaliados: comprimento do fruto em mm (CF); diâmetro do fruto em mm (DF); peso do

¹DET - Universidade Federal de Viçosa; e-mail: vinicius.s.santos@ufv.br

³Agradecimento a Capes e FAPEMIG pelo apoio financeiro.

²Embrapa Amazônia Oriental.

⁴Agradecimento a EMBRAPA e ao CNPq pelo apoio financeiro.

fruto em g (PF); peso de polpa em g (PP); peso de casca em g (PC); espessura de casca em mm (EC); peso de fibra em g (PFI); peso médio de todas as sementes por fruto em g (PS) e número de sementes (NS). Os dados foram inicialmente analisados por meio da metodologia de modelos mistos REML/BLUP (máxima verossimilhança restrita/melhor predição linear não viesada), sendo adotado o seguinte modelo matemático (RESENDE, 2007a):

$$y = Xb + Za + e; \quad (1)$$

em que y é o vetor de dados, b é o escalar referente à média geral (efeito fixo), a é o vetor dos efeitos genotípicos (assumidos como aleatórios) e e é o vetor de erros aleatórios. X e Z são as matrizes de incidência para b e a , respectivamente.

No método tradicional, os componentes principais de um conjunto de variáveis correlacionadas são função lineares das variáveis originais, sendo essas novas variáveis não correlacionadas, e que explicam o máximo da variação original. A matriz de covariâncias das variáveis originais é dada por

$$\Sigma = \Lambda\Lambda' = VD_{\alpha}V'; \quad (2)$$

em que D_{α} é a matriz diagonal dos autovalores, V é a matriz dos autovetores e $\Lambda = V(D_{\alpha})^{1/2}$. A partir dessa composição da matriz de covariância, os componentes principais são dados por $PC_i = v_i y$, em que y é o vetor das variáveis originais, v_i é o autovetor i associado ao autovalor i , com $v'v_i = 1$. Assim, os pesos das variáveis em cada componente principal é dado pelos elementos do autovetor correspondente. A variância de cada PC é dada pelo correspondente autovalor. Os primeiros PCs associados aos primeiros autovalores ordenados e respectivos autovetores explicam a maior parte da variação total. Resende (2007a) sugere que escolha-se o número de componentes que expliquem a porção da variação total superior a 80%. As estimativas dos autovalores são obtidas por $\det(R - \alpha_i I) = 0$ e dos autovetores são obtidas pela solução do sistema $(R - \alpha_i I)v_i = \Phi$, em que R é a matriz de correlações entre as variáveis, α_i refere-se ao autovalor i , v_i é o autovetor i e Φ é um vetor nulo.

O modelo (1) pode ser reescrito da seguinte forma, visando a PCAM:

$$y = Xb + Z^*a^* + \varepsilon = Xb + Z(Q \otimes I_g)(Q^{-1} \otimes I_g)a + \varepsilon; \quad (3)$$

em que $Q = V_m$ e $a_j^* = Q'a_j$. Os valores genéticos do indivíduo j para as variáveis originais é dado por $a_j^* = Q'a_j$. I_g é a matriz identidade com ordem igual ao número g de genótipos. Sob esse modelo, a matriz de covariância genética é dada de forma semelhante à (2). D_{α} é a matriz diagonal dos m autovalores e V é a matriz dos autovetores. Escolhendo-se V e D_{α} referentes apenas à dimensão m , esse modelo misto é reduzido e ajusta somente os m primeiros componentes principais. O método REML é então aplicado para estimação nesse modelo. Mais detalhes em Resende (2007a). Para a estimação dos componentes de variância/predição dos valores genéticos e a Análise de Componentes Principais sob modelo misto (PCAM) foi utilizado

o software Selegen-Reml/Blup, versão 2009 (RESENDE, 2007b).

Para o estudo de diversidade genética entre os clones de cupuaçuzeiro, foi realizada a análise de agrupamento via os métodos hierárquico aglomerativo de ligação média não ponderada UPGMA e de otimização de Tocher, baseados na matriz de distâncias euclidiana média, utilizando como variáveis os escores obtidos por meio da análise de componentes principais sob modelos mistos (PCAM). A adequação do agrupamento no método hierárquico foi verificada por meio do coeficiente de correlação cofenética (CCC), em que o valor obtido foi avaliado pelo teste de Mantel ao nível de significância de 5% e com 1000 permutações. Para a escolha do número de grupos foi utilizado o critério de Mojena, com o valor de $k=1,25$. Para essas análises foi utilizado o programa Genes, versão 2013.5.1 (CRUZ, 2006).

3 Resultados e Discussão

Na Tabela 1 encontram-se os autovalores e os autovetores das 9 características avaliadas em cupuaçuzeiro. Observa-se que os três primeiros componentes principais explicaram cerca de 80,15% da variação total. Araújo et al. (2002) avaliando os mesmos caracteres e algumas relações entre eles, obtiveram 76% da variação total explicada por dois componentes baseados em valores fenotípicos. A análise dos dados obtidos com o emprego da distância euclidiana média possibilitou a identificação da maior distância entre os clones 9 e 25 (2,62), enquanto que os clones 6 e 18 (0,06) foram os mais similares.

Tabela 1: Estimativas dos Autovalores e Autovetores obtidos a partir da análise de 9 caracteres de cupuaçuzeiro, Tomé-Açu, Pará, 2014.

Variáveis	1º Componente	2º Componente	3º Componente
CFR	0,3288	-0,4281	-0,0378
DFR	0,2711	0,5387	0,1175
PFR	0,4761	0,0314	0,1162
PPO	0,3728	0,3117	0,3088
PCA	0,4459	-0,1609	-0,1889
ECA	0,3203	-0,3038	-0,3976
PFI	0,3291	-0,0092	-0,1491
PSE	0,1643	-0,187	0,72
NS	-0,1265	-0,5246	0,3761
Autovalor	4,2498	1,5958	1,368
PVE (%)	47,22	17,73	15,2
PVE Acumulada (%)	47,22	64,95	80,15

PVE (%): Proporção da variabilidade explicada pelos Autovalores; Comprimento do Fruto em mm (CF); Diâmetro do Fruto em mm (DF); Peso do Fruto em g (PF); Peso de Polpa em g (PP); Peso de Casca em g (PC); Espessura de Casca em mm (EC); Peso de Fibra em g (PFI), Peso médio de todas as sementes por fruto em g (PS) e Número de Sementes (NS).

Com base nos resultados de agrupamento pelo método de Tocher, os 50 clones de cupuaçuzeiro

formaram 7 grupos, sendo que o grupo I foi composto por 22 clones, o grupo II por 12, o grupo III por 5 clones, o grupo IV por 7 clones e o grupo V por dois clones. Os grupos VI e VII foram formados por clones isolados (Tabela 2). O dendrograma obtido pelo método hierárquico aglo-

Tabela 2: Grupos formados pelo método de Tocher a partir das distâncias euclidianas médias estimadas por meio dos três primeiros componentes para os 50 clones de cupuaçuzeiro, Tomé-Açu - Pará, 2014.

Grupos	Clones
I	6;18;22;1;4;20;7;49;47;30;2;38;23;41;50;37;44;14;5;29;40;3
II	32;45;46;21;28;19;11;27;15;10;16;43
III	8;34;17;42;25
IV	35;39;48;12;26;33;36
V	24;31
VI	13
VII	9

merativo de ligação média não ponderada (UPGMA) com base na distância euclidiana média é mostrado na Figura 1. O coeficiente de correlação cofenética (CCC) foi igual a 0,69, sendo significativo (valor-p < 0,001) pelo teste de Mantel, indicando que o agrupamento foi adequado. De acordo com o critério de Mojena, observa-se a formação de seis grupos. O grupo I foi

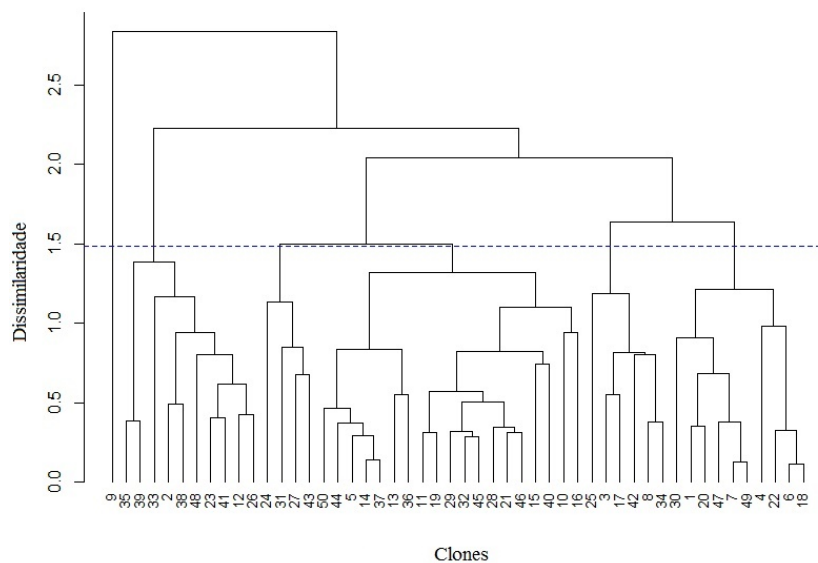


Figura 1: Dendrograma da divergência genética entre 50 clones de cupuaçuzeiro, gerado pelo método de agrupamento UPGMA, Pará, 2014.

formado por 10 clones. Esses mesmos clones pertencem ao grupo I formado pelo método de otimização de Tocher. O grupo II foi composto por 19 clones. Esses mesmos clones pertencem aos grupos I e II formados pelo método de Tocher, com exceção dos clones 36 e 13, pertencentes aos grupos IV e VI, respectivamente. O grupo III foi formado pelos clones: 34, 8, 42, 17, 3 e 25, todos pertencentes ao grupo III do método de Tocher, exceto o clone 3, pertencente ao grupo

I. O grupo IV foi composto por 10 clones e o grupo V foi formado pelos clones: 43, 27, 31 e 24. Em ambos os métodos, o clone 9 formou grupo isolado. Utilizando o método de Tocher, Maia et al. (2011b), ao avaliarem 8 clones de cupuaçuzeiro, obtiveram quatro grupos. Entretanto, na avaliação de 36 progênies de cupuaçuzeiro, Maia et al. (2011a) obtiveram grupos menores, dois e três, a partir das distâncias quadradas euclidianas e de Mahalanobis, respectivamente.

4 Conclusões

Existe variabilidade genética entre os clones de cupuaçuzeiro estudados, sendo os clones 9 e 25 mais divergentes. O método de agrupamento hierárquico UPGMA e o não hierárquico de Tocher foram concordantes na formação dos grupos, sendo que o método UPGMA formou seis e o método de Tocher formou sete grupos.

Referências

- [1] ALVES, R.M.; RESENDE, M.D.V.; BANDEIRA, B.S.; PINHEIRO, T.M.; FARIAS, D.C.R. Avaliação e seleção de progênies de cupuaçuzeiro (*Theobroma grandiflorum*), em Belém, Pará. **Revista Brasileira de Fruticultura**. Jaboticabal. v.32, n. 1, p. 204-212, 2010.
- [2] ARAÚJO, D.G. de; CARVALHO, S.P.; ALVES, R.M. Divergência genética entre clones de cupuaçuzeiro (*Theobroma grandiflorum* Willd. ex Spreng. Schum.). **Ciência e Agro-tecnologia**. Lavras. v. 26, n. 1, p. 13-21, 2002.
- [3] CRUZ, C.D. **Programa Genes: Biometria**. Viçosa: Editora UFV, 2006, 382 p.
- [4] MAIA, M.C.C.; RESENDE, M.D.V.; OLIVEIRA, L.C.; ALVES, R.M.; SILVA FILHO, J.L.; ROCHA, M.M.; CAVALCANTE, J.J.V.; RONCATTO, G. Análise genética de famílias de meios-irmãos de cupuaçuzeiro. **Pesquisa Florestal Brasileira**. Colombo. v. 31, n. 66, p. 123-130, 2011a.
- [5] MAIA, M.C.C.; RESENDE, M.D.V.; OLIVEIRA, L.C.; ÁLVARES, V.S.; MACIEL, V.T.; LIMA, A.C. Seleção de clones experimentais de cupuaçu para características agroindustriais via modelos mistos. **Revista Agro@mbiente On-line**, Boa Vista, v.5, n.1, p.35-43, 2011b.
- [6] RESENDE, M.D.V. **Matemática e estatística na análise de experimentos e no melhoramento genético**. Colombo: Embrapa Florestas, 2007a. 561 p.
- [7] RESENDE, M.D.V. **SelegenReml/Blup: Sistema Estatístico e Seleção Genética Computadorizada via Modelos Lineares Mistos**. Colombo: Embrapa Florestas, 2007b. 360 p.