

# DIFERENTES COEFICIENTES DE SIMILARIDADE PARA ANÁLISE DE AGRUPAMENTO DE GENÓTIPOS DE *Coffea canephora* COM BASE EM MARCADORES MOLECULARES

Maria Amélia Gava FERRÃO<sup>1</sup>; Aymbiré Francisco Almeida da FONSECA<sup>2</sup>; Elaine Manelli RIVA-SOUZA<sup>3</sup>; Romário Gava FERRÃO<sup>4</sup>

<sup>1</sup>Pesquisadora Embrapa Café/Incaper, Bolsista CNPq, E-mail: mferrao@incaper.es.gov.br; <sup>2</sup>Pesquisador Embrapa/Incaper; <sup>3</sup>Bolsista CBP&D-Café/Incaper; <sup>4</sup>Pesquisador Incaper

## Resumo:

O objetivo deste trabalho foi comparar o agrupamento de 80 genótipos de *C. canephora* com 15 diferentes coeficientes de similaridade, obtidos por meio de marcadores moleculares RAPD. Os coeficientes de similaridade analisados foram: Jaccard, Nei e Li, Ochiai, Índice II, Coincidência simples, Rogers e Tanimoto, Sokal e Sneath, Yule, Índice I, Haman, Ochiai II, Índice III, Russel e Rao, Baroni et al. e Sokal. As estimativas de similaridade ( $S_{g_{ii}}$ ) entre todos os pares de genótipos ( $ii'$ ) foram convertidas em dissimilaridade por meio do complemento aritmético dos coeficientes utilizados ( $D_{g_{ii}} = 1 - S_{g_{ii}}$ ). Para a análise de agrupamento dos genótipos foi empregado o método de otimização de Tocher. Verificaram-se diferenças quanto ao número de grupos formados com a utilização dos diferentes coeficientes de similaridade, variando de quatro a 43. Houve concordância entre os coeficientes de Jaccard, Nei e Li, Ochiai, Índice II, Coincidência simples, Rogers e Tanimoto, Sokal e Sneath, Yule, Índice I, Haman, Ochiai II, Índice III e Baroni et al. quanto ao agrupamento dos genótipos mais dissimilares e mais similares. Os coeficientes de Russel e Rao e Sokal apresentaram dados discrepantes e sugere-se que não sejam utilizados. Os resultados em conjunto caracterizam a necessidade de estudos complementares para acurácia na definição do (s) coeficiente (s) para a análise de dados moleculares obtidos por meio de marcadores dominantes RAPD em *C. canephora*.

Palavras-chave: *Coffea*, coeficientes de similaridade, divergência genética, marcadores RAPD.

## DIFFERENT COEFFICIENTS OF SIMILARITY IN THE GROUPING OF GENOTYPES OF *Coffea canephora* BY MEANS OF MOLECULAR MARKERS

### Abstract:

The objective of this work was to compare the grouping of genotypes of *C. canephora* by using 15 different coefficients of similarity, obtained by means of RAPD molecular markers. The coefficients of similarity analyzed were: Jaccard, Nei and Li, Ochiai, Índice II, Simple coincidence, Rogers and Tanimoto, Sokal and Sneath, Yule, Índice I, Haman, Ochiai II, Índice III, Russel and Rao, Baroni et al. and Sokal. To analyze the grouping of the genotypes the method of optimization of Tocher was used. Differences in the number of groups formed using the different coefficients of similarity, varying from four to 43, were verified. There was agreement between the coefficients of Jaccard, Nei and Li, Ochiai, Índice II, Simple coincidence, Rogers and Tanimoto, Sokal and Sneath, Yule, Índice I, Haman, Ochiai II, Índice III and Baroni et al. in grouping the genotypes most dissimilar and most similar; the coefficients of Russel and Rao and Sokal presented data discrepancies thus their use is not suggested. The results demonstrate the necessity of complementary studies for a precise definition of the best coefficient for analysis of molecular data obtained by dominant RAPD markers in *C. canephora*.

Key words: *Coffea*, coefficients of similarity, genetic divergence, RAPD markers.

## Introdução

A espécie *Coffea canephora* é a mais cultivada no Espírito Santo e o Estado destaca-se como o principal produtor de café Conilon do Brasil.

O Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper) possui um Banco Ativo de Germoplasma (BAG) da espécie com 375 acessos e vem se dedicando a pesquisas com esta cultura a partir de 1985, abrangendo diversas áreas do conhecimento, entre elas, recursos genéticos, melhoramento genético vegetal e biologia molecular (Ferrão et al., 2004).

Marcadores morfológicos e moleculares vêm sendo empregados para estudar a diversidade genética em bancos de germoplasma, concomitantemente com a utilização de diferentes metodologias de análises multivariada. O método de otimização de Tocher permite o estabelecimento de grupos considerando-se a existência de homogeneidade dentro do grupo e heterogeneidade entre grupos (Cruz, 2001).

Antes do emprego de qualquer método, uma matriz com valores de similaridade ou distâncias genéticas deve ser obtida entre os genótipos (Meyer et al, 2004) por meio de coeficientes de similaridade. Tratando-se da análise com dados

oriundos de marcadores moleculares, diferentes abordagens podem ser consideradas, implicando na escolha de diferentes coeficientes.

Na literatura, observa-se que distintos coeficientes são aplicados com o mesmo propósito, assim como um determinado coeficiente sendo utilizado com culturas e objetivos variados, sem a devida justificativa da escolha pelos autores (Emygdio et al., 2003). Nos estudos de diversidade genética com o gênero *Coffea* sp., alguns coeficientes diferentes têm sido utilizados, sem contudo conhecimento de pesquisas comparativas sobre a aplicação de cada um deles.

A similaridade e dissimilaridade genética com base em marcadores moleculares é avaliada com base numa matriz de dados originais, onde a presença de determinada banda de DNA entre dois genótipos é codificada como 1 e a ausência da banda, como 0. Assim, comparando-se genótipos dois a dois, as seguintes situações podem ser apresentadas: a = 1:1 (presença de banda nos dois genótipos), b = 1:0 (presença de banda no genótipo 1 e ausência no genótipo 2), c = 0:1 (ausência de banda no genótipo 1 e presença no genótipo 2) e d = 0:0 (ausência de bandas nos dois genótipos).

O objetivo do presente trabalho foi comparar diferentes coeficientes de similaridade na análise de agrupamento de genótipos de *Coffea canephora* com base em marcadores RAPD.

## Material e Métodos

Neste estudo utilizaram-se folhas de 80 genótipos da espécie *Coffea canephora* Pierre ex Froenher, mantidos no Banco Ativo de Germoplasma (BAG) do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural – Incaper, situado na Fazenda Experimental de Marilândia, município de Marilândia-ES. O trabalho foi desenvolvido no Laboratório de Biologia Molecular do Incaper, localizado no Centro Regional de Desenvolvimento Rural Centro Serrano, município de Venda Nova do Imigrante-ES.

Para a extração do DNA, adotou-se o protocolo proposto por Doyle e Doyle (1990), com algumas modificações. As amostras de DNA foram amplificadas em termociclador GeneAmp PCR System 9700, pela técnica de RAPD (Random Amplified Polymorphic DNA) de acordo com Williams et al. (1990), utilizando-se um volume de 25 µL, contendo 25 ng de DNA, 100 µM de cada um dos desoxinucleotídeos (dATP, dCTP, dGTP e dTTP); MgCl<sub>2</sub> 2,0 mM; Tris-HCl 10 mM (pH 8,3); KCl 50 mM; 0,4 µM de um primer decâmero e uma unidade da enzima *Taq* polimerase. As condições de amplificação foram as seguintes: desnaturação do DNA a 94°C, por 15 segundos, ligação do primer ao DNA molde a 35°C, por 30 segundos, e extensão da fita de DNA a 72°C, por um minuto. Após 40 ciclos, foi efetuada uma última etapa de extensão a 72°C, por sete minutos. Os produtos de amplificação foram separados por eletroforese horizontal utilizando-se gel de agarose 1,2%, imerso em tampão TBE (Tris-Borato 90 Mm, EDTA 1 mM, pH 8,0), a 100 volts durante duas horas e 45 minutos. O gel foi corado com brometo de etídeo (0,5 µg/µL) por 30 minutos e descorado em água destilada durante 30 minutos. Os fragmentos de DNA foram visualizados com o sistema de fotodocumentação Eagle Eye II (Stratagene).

A análise dos dados moleculares foi baseada na codificação 1 para presença e 0 para ausência de bandas de mesmo peso molecular no gel de agarose.

Os coeficientes de similaridade analisados foram: Jaccard, Nei e Li, Ochiai, Índice II, Coincidência simples, Rogers e Tanimoto, Sokal e Sneath, Yule, Índice I, Haman, Ochiai II, Índice III, Russel e Rao, Baroni et al. e Sokal. As estimativas de similaridade ( $S_{g_{ii}}$ ) entre todos os pares de genótipos (ii') foram convertidas em dissimilaridade por meio do complemento aritmético dos coeficientes utilizados ( $D_{g_{ii}} = 1 - S_{g_{ii}}$ ) (Cruz e Carneiro, 2003). Para a análise de agrupamento dos genótipos foi empregado o método de otimização de Tocher. As análises estatísticas foram realizadas com auxílio do software Genes (Cruz, 2001), versão 2006.4.1.

## Resultados e Discussão

O número de grupos formados por meio do método de otimização de Tocher com base nos 15 coeficientes foi bastante variável, partindo da formação de quatro até 43 grupos (Tabela 1). A título de ilustração foram apresentados na Tabela 2 os resultados dos agrupamentos observados para os coeficientes Sokal, Russel e Rao, Coincidência simples, Jaccard e Índice II.

Os coeficientes de similaridade de Jaccard, Nei e Li, Ochiai e Índice II não consideram a opção ausência (0:0) de banda como sinônimo de similaridade entre os genótipos em suas expressões. Os resultados demonstraram não haver grandes discrepâncias na estrutura do agrupamento dos genótipos utilizando-se estes coeficientes. Jaccard e Nei e Li formaram exatamente os mesmos 15 grupos, permanecendo 68,75% dos genótipos no grupo 1. Para Ochiai e Índice II, algumas alterações na ordem de agrupamento foram verificadas. Ochiai formou 18 grupos, com 61,25% dos genótipos no grupo 1 e o Índice II permitiu a formação de 19 grupos, mantendo-se 56,25% dos genótipos no grupo 1. De maneira geral, estes coeficientes permitiram a identificação dos genótipos mais similares sempre alocados no grupo 1 e apontaram os mesmos genótipos considerados mais dissimilares. Estes resultados concordam com Meyer et al (2004) que verificaram em estudos de divergência genética em milho, por meio de marcadores moleculares dominantes RAPD, que os coeficientes de Jaccard e Ochiai apresentaram resultados próximos, pois todos excluem as ocorrências negativas (0:0). Estes autores sugerem ainda, que os coeficientes que não consideram as ausências de fragmentos como evidências de homologia são os mais indicados para a obtenção da matriz de dissimilaridade para marcadores dominantes.

Tabela 1 - Coeficientes de similaridade utilizados para obtenção da matriz de divergência genética entre 80 genótipos de *C. canephora* e número de grupos formados, com base em marcadores moleculares dominantes RAPD.

Coeficiente	Expressão da similaridade <sup>(1)</sup>	Número de grupos
Jaccard	$a/(a+b+c)$	15
Nei e Li	$2a/(2a+b+c)$	15
Ochiai	$a/\sqrt{(a+b)(a+c)}$	18
Índice II	$0,5[a/(a+b) + a/(a+c)]$	19
Coincidência simples	$(a+d)/(a+b+c+d)$	6
Rogers e Tanimoto	$(a+d)/[a+2(b+c)+d]$	6
Sokal e Sneath	$2(a+d)/[2(a+d)+b+c]$	7
Yule	$(ab-bc)/(ad+bc)$	4
Índice I	$(ad-bc)/(ad+bc)$	4
Haman	$[(a+d)-(b+c)]/(a+b+c+d)$	6
Ochiai II	$ad/\sqrt{(a+b)(a+c)(d+b)(d+c)}$	4
Índice III	$0,25[a/(a+b) + a/(a+c) + d/(d+b) + d/(d+c)]$	4
Baroni et al.	$[a + \sqrt{(ad)}]/[a+b+c + \sqrt{(ad)}]$	10
Russel e Rao	$a/(a+b+c+d)$	20
Sokal	$\sqrt{[(b+c)/(a+b+c+d)]}$	43

<sup>(1)</sup> a=1:1; b=1:0; c=0:1 e d=0:0.

Com relação aos coeficientes que consideram na expressão a ausência de bandas para os dois genótipos, ou seja 0:0=d, verificou-se por meio do coeficiente de coincidência simples que seis grupos foram formados, permanecendo o grupo 1 com 92,5% dos genótipos. Observação semelhante foi feita para os coeficientes de Rogers e Tanimoto e Haman, porém diferindo na ordem dos genótipos dentro do grupo 1. O índice de Sokal e Sneath contribuíram para a formação de sete grupos, sendo o primeiro grupo composto por 91,25% do total dos genótipos. Os coeficientes de Yule, Índice I e Índice III apresentaram resultados coincidentes, com a composição de quatro grupos. Os grupos 1 de Yule, Índice I e Índice III foram compostos por 95% dos genótipos, ordenados com algumas alterações. Com o coeficiente de Ochiai II também houve a formação de quatro grupos, porém com 96,25% dos genótipos no grupo 1. Embora estes coeficientes consideram a ausência de ocorrência de bandas (0:0) como similaridade entre os genótipos em suas expressões, seguem o mesmo padrão de genótipos mais similares e genótipos mais dissimilares observados para os coeficientes que desconsideram a ocorrência 0:0 (Jaccard, Nei e Li, Ochiai e Índice II). Contudo, pelo menor número de grupos formados, verificou-se que estes coeficientes não são tão eficazes para discriminar os genótipos.

O coeficiente de Baroni et al. formou 10 grupos com 76,25% dos genótipos concentrados no grupo 1. Seu padrão de agrupamento e a ordem dos genótipos, bem como a classificação de genótipos mais similares e mais dissimilares foi semelhante aos coeficientes que não consideram as ocorrências 0:0 (Jaccard, Nei e Li, Ochiai e Índice II) em suas expressões.

Considerando-se Russel e Rao, 20 grupos foram formados, sendo o Grupo 1 composto por 72,5% do total de genótipos estudados. Este coeficiente apresentou agrupamento e ordem de genótipos bastante dissimilar dos demais coeficientes estudados, sugerindo-se que não seja recomendado para o estudo da divergência em *C. canephora*. Meyer et al (2004) sugeriram que o coeficiente de Russel e Rao não fosse utilizado em milho, com dados RAPD, pois observaram resultados bastante distintos dos demais coeficientes estudados, considerando que o coeficiente exclui as ocorrências negativas (0:0) do numerador e inclui-as no denominador da expressão.

O coeficiente de Sokal permitiu o agrupamento dos genótipos em 43 grupos. Verificou-se que a estrutura do agrupamento foi também bastante contraditória aos demais coeficientes estudados, uma vez que Sokal agrupou os genótipos tidos como mais similares pelos outros coeficientes, como os mais dissimilares e vice-versa. Portanto, sugere-se que este coeficiente não seja recomendado para a análise de dados RAPD em *C. canephora*. A exemplo de Russel e Rao, este coeficiente exclui as ocorrências negativas (0:0) do numerador e inclui-as no denominador da expressão. As ocorrências positivas (1:1) estão inclusas apenas no denominador.

Tabela 2 - Agrupamento de 80 genótipos de *C. canephora* por meio do método de otimização de Tocher, mediante a utilização de cinco coeficientes de similaridade.

Grupo	Sokal	Russel e Rao	CS	Jaccard	Índice II
1	15, 25, 41, 7, 56, 69, 45, 61, 80, 30, 78, 37, 9, 60, 6, 20, 33, 43, 16, 64, 49, 17, 31, 35	50, 51, 34, 26, 36, 21, 55, 72, 73, 54, 52, 62, 64, 33, 24, 29, 38, 37, 30, 35, 11, 12, 57, 32, 23, 14, 6, 25, 67, 65, 58, 19, 74, 66, 44, 10, 69, 71, 60, 70, 63, 27, 56, 53, 40, 41, 68, 59, 43, 47, 39, 2, 42, 22, 28, 4, 13, 48	11, 12, 18, 57, 44, 51, 50, 54, 19, 39, 55, 72, 73, 14, 26, 74, 62, 34, 21, 36, 29, 23, 30, 24, 32, 22, 38, 10, 70, 66, 63, 48, 65, 27, 42, 71, 4, 40, 53, 67, 68, 13, 5, 79, 59, 2, 1, 64, 58, 33, 28, 35, 37, 3, 43, 47, 77, 75, 46, 6, 17, 69, 16, 15, 52, 76, 31, 8, 7, 60, 9, 20, 49, 61	11, 12, 18, 57, 44, 51, 50, 54, 19, 39, 55, 72, 73, 14, 26, 62, 74, 34, 21, 36, 29, 23, 30, 24, 32, 22, 38, 10, 70, 66, 63, 48, 65, 27, 42, 71, 4, 40, 53, 67, 68, 13, 5, 79, 59, 2, 1, 64, 58, 33, 28, 35, 37, 3, 43, 47, 77, 75, 46, 6, 17, 69, 16, 15, 76, 52, 31, 8, 7, 60, 9, 20, 49	11, 12, 57, 44, 51, 54, 55, 73, 72, 26, 62, 66, 65, 64, 74, 36, 21, 34, 14, 23, 24, 29, 30, 39, 19, 18, 32, 33, 38, 50, 10, 67, 70, 63, 71, 52, 58, 22, 53, 48, 37, 42, 27, 35, 40
2	5, 58	7, 8, 3	78, 80	7, 8, 1, 16, 5, 3	7, 8, 4, 5, 3, 1, 16, 13
3	8, 40, 52	45, 46	45	60, 61	68, 76, 69, 77
4	2, 18	18	56	45, 46, 49	59, 60, 61
5	38, 77, 47	49	41	75, 79, 17	47, 49, 46, 6
6	44, 76, 28	80	25	76, 77	75, 79
7	1, 46	20		80	2, 15
8	55, 75	17		9	78
9	3, 65	77		15	20
10	23, 67	61		20	43
11	59, 71	76		78	28
12	29, 66	5		31	80
13	53	9		56	17
14	13	1		25	9
15	79	16		41	31
16	57	79			56
17	4	78			41
18	42	31			45
19	11	15			25
20	12	75			
21	50				
22	19				
23	68				
24	32				
25	70				
26	21				
27	74				
28	10				
29	48				
30	34				
31	24				
32	22				
33	14				
34	54				
35	27				
36	39				
37	72				
38	36				
39	63				
40	73				
41	51				
42	26				
43	62				

## Conclusões

- O número de grupos formados por meio do método de otimização de Tocher com base nos 15 coeficientes foi bastante variável, partindo da formação de quatro (Yule) até 43 grupos (Sokal).
- Houve concordância entre os coeficientes de Jaccard, Nei e Li, Ochiai, Índice II, Coincidência simples, Rogers e Tanimoto, Sokal e Sneath, Yule, Índice I, Haman, Ochiai II, Índice III e Baroni et al. quanto ao agrupamento dos genótipos mais dissimilares e mais similares.
- Os coeficientes de Russel e Rao e Sokal apresentaram dados discrepantes e sugere-se que não sejam utilizados.
- Os resultados em conjunto caracterizam a necessidade de estudos complementares para se indicar com acurácia o melhor coeficiente para a análise de dados moleculares obtidos por meio de marcadores dominantes RAPD na espécie *C. canephora*.

## Agradecimento

Os autores agradecem ao Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D Café) pelo apoio financeiro e pela concessão de bolsas de desenvolvimento, e ao Dr. Mark Kulig por sua participação neste trabalho.

## Referências Bibliográficas

- Cruz, C.D. (2001) *Programa Genes*: versão Windows. Aplicativo computacional em genética e estatística. Viçosa: UFV, 648p.
- Cruz, C.D., Carneiro, P.C.S. (2003) *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: UFV, 585p.
- Doyle, J.J., Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus*, 12:13-15.
- Emygdio, B.M., Antunes, I.F., Choer, E., Nedel, J.L. (2000) Eficiência de coeficientes de similaridade em genótipos de feijão mediante marcadores RAPD. *Pesq. Agropec. Bras.*, 38(2):243-250.
- Ferrão, R.G., Fonseca, A.F.A. da, Ferrão, M.A.G., De Muner, L.H., Verdin Filho, A.C., Volpi, P.S., Marques, E.M.G., Zucatei, F. (2004) *Café conilon*: técnicas de produção com variedades melhoradas. Vitória, ES: Incaper, 60p. (Circular Técnica, 03-I)
- Meyer, A.das, Garcia, A.A.F., Souza, A.P. de, Souza Júnior, C.L.de. (2004) Comparison of similarity coefficients used for cluster analysis with dominant markers in maize (*Zea mays* L.). *Genetics and Molecular Biology*, 27(1):83-91.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research*, 18(22):6531-6535.

