

# Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD

Nair Helena Castro Arriel<sup>(1)</sup>, Antônio Orlando Di Mauro<sup>(2)</sup>, Sônia Marli Zingaretti Di Mauro<sup>(2)</sup>, Olaf Andreas Bakke<sup>(3)</sup>, Sandra Helena Unêda-Trevisoli<sup>(2)</sup>, Marcelo Marchi Costa<sup>(2)</sup>, Andréa Capeloto<sup>(2)</sup> e Amanda Roberta Corrado<sup>(2)</sup>

<sup>(1)</sup>Embrapa Algodão, Rua Osvaldo Cruz, nº 1143, CEP 58107-720 Campina Grande, PB. E-mail: nair@cnpa.embrapa.br <sup>(2)</sup>Universidade Estadual Paulista, Fac. de Ciências Agrárias e Veterinárias, Via de Acesso Prof. Paulo Donato Castellane, s/nº, CEP 14884-900 Jaboticabal, SP. E-mail: orlando@fcav.unesp.br, zingara@fcav.unesp.br, mmarchi@hotmail.com, shutrevi@fcav.unesp.br, andcapeloto@aol.com, acorrado@bol.com.br <sup>(3)</sup>Universidade Federal de Campina Grande, Campus de Patos/CSTR, CEP 58700-970 Patos, PB. E-mail: obakke@cstr.ufcg.edu.br

**Resumo** – O objetivo deste trabalho foi comparar diferentes técnicas multivariadas na caracterização de 35 genótipos de gergelim mediante 769 marcadores RAPD. As distâncias genéticas foram obtidas pelo complemento aritmético do coeficiente de Jaccard e agrupadas pelos métodos hierárquicos do vizinho mais próximo, do vizinho mais distante, das médias aritméticas não ponderadas (UPGMA), do método de otimização de Tocher e análises de coordenadas principais. O agrupamento dos genótipos foi alterado em função dos diferentes métodos usados. Adotando-se a mesma distância genética (0,36) como valor de corte, diferenciaram-se quatro grupos no método do vizinho mais próximo, 13 para o vizinho mais distante, 11 no UPGMA e quatro no Tocher. Entre os métodos hierárquicos, o UPGMA apresentou o melhor ajuste das distâncias originais e estimadas (CCC = 0,89). As análises das coordenadas principais confirmaram a baixa diversidade existente entre os genótipos. A maior divergência ocorreu entre as cultivares Seridó 1 e Arawaca 4, e a menor, entre os genótipos VCR-101 e GP-3314. As três primeiras coordenadas principais contabilizaram 35,13% do total da variabilidade, e 18 autovalores foram necessários para explicar 81% da variação genética. Os métodos UPGMA, de otimização de Tocher, e as análises de coordenadas principais são complementares na formação dos grupos.

Termos para indexação: *Sesamum indicum*, análise de agrupamento, coordenadas principais.

## Multivariate techniques for the determination of genetic diversity in sesame using RAPD markers

**Abstract** – The objective of this work was to compare different multivariate techniques in the characterization of 35 sesame genotypes using 769 RAPD makers. Genetic distances were obtained from the arithmetic complement of the Jaccard coefficient, and were evaluated by single linkage, complete linkage and unweighted arithmetic average (UPGMA) agglomerative methods, Tocher optimization method and principal coordinate analysis. The clustering structure was altered by the different methods. Adopting the same genetic distance (0.36) as value of cut, four groups were discriminated in single linkage, 13 in complete linkage and 11 in UPGMA; the Tocher's optimization methods formed four groups. Among the hierarchical clustering methods, UPGMA showed the best adjust for estimated and originals distances (CCC = 0.89). Principal coordinates analyses confirmed the low diversity among genotypes. The highest divergence occurred between Seridó 1 and Arawaca 4 cultivars, and the minor, between VCR-101 and GP-3314 genotypes. The three first principal coordinates responded for 35.13% of the variability, and 18 autovalues were necessary to explain 81% of the genetic variation. UPGMA, Tocher's optimization and the principal coordinates analyses are complementary in the clusters formation.

Index terms: *Sesamum indicum*, clustering analyses, principal coordinates.

### Introdução

O gergelim (*Sesamum indicum* L.) é a oleaginosa de uso mais antigo, originária dos continentes africano e asiático, e no Brasil vem sendo explorado há mais de 60 anos. Suas sementes apresentam considerável teor

de proteínas e sais minerais e fornecem um óleo rico em ácidos graxos de elevada estabilidade química. Em razão de sua facilidade de cultivo e adaptabilidade às zonas áridas e semi-áridas do Nordeste, desde 1986 a Embrapa Algodão, com outras instituições de pesquisa, vem desenvolvendo trabalhos de melhoramento genéti-

co que têm contribuído para o estudo da adaptabilidade e da estabilidade de genótipos, pela obtenção de cultivares altamente produtivas e adaptadas.

A caracterização da diversidade genética do gergelim tem sido feita com base em marcadores morfológicos e agronômicos. Manivannan & Nadarajan (1996) constataram que a altura da planta, número de ramificações, rendimento de sementes e cápsulas por planta foram as características que mais contribuíram para a caracterização de 52 acessos.

Singh et al. (1997) observaram, também, que a altura da planta, o rendimento de sementes e o conteúdo de óleo foram os que mais contribuíram para variabilidade entre 33 genótipos.

Swain & Dikshit (1997) identificaram o conteúdo de óleo, massa de mil sementes, comprimento da cápsula e início de floração como as características que mais contribuíram para explicação da variabilidade entre 40 genótipos de gergelim.

Apesar de ser descrito como planta autógama, o gergelim apresenta alta manifestação heterótica em combinações híbridas, tanto para os caracteres de produção quanto para caracteres vegetativos (Arriel et al., 2001). Nesse sentido, o conhecimento da diversidade, baseado em informações mais precisas, como as geradas pelos marcadores moleculares, além de direcionar as estratégias de cruzamentos específicos, maximizando os ganhos genéticos nos ciclos de seleção, permite otimizar procedimentos para a conservação de recursos genéticos, pela avaliação de redundâncias e deficiências da base genética disponível (Dias, 1998).

Dentro da classe de marcadores moleculares, baseados na técnica de reação em cadeia da polimerase (PCR), destaca-se o polimorfismo de DNA amplificado ao acaso (RAPD). Nos últimos anos, vários estudos têm demonstrado a eficiência dos marcadores RAPD na caracterização da variabilidade genética (Pereira & Kerr, 2001; Bered et al., 2002; Emygdio et al., 2003; Sera et al., 2003). Estudos dessa natureza com o gergelim são restritos; padrões enzimáticos foram usados por Isshiki & Umezaki (1997), em 68 acessos provenientes de diferentes países, e os autores constataram a estreita base genética do gergelim cultivado.

Bhat et al. (1999) constataram uma alta divergência genética em germoplasma de gergelim indiano e exótico, a partir de marcadores RAPD.

Na análise de dados moleculares, métodos multivariados são empregados, como os de agrupamento (hierarquização e otimização) e os de ordenação

(componentes principais e coordenadas principais). A análise de agrupamento procura discriminar geneticamente os indivíduos, e permite separá-los em grupos pela análise de um conjunto de características inerentes a cada indivíduo, agrupando-os por algum critério de classificação, de forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (Cruz & Carneiro, 2003). Basicamente, os métodos hierárquicos são utilizados na caracterização da diversidade genética, por meio de marcadores RAPD (Pereira & Kerr, 2001; Sanz-Cortés et al., 2001; Oliveira et al., 2002; Costa et al., 2003; Emygdio et al., 2003).

De maneira geral, ao se fazer o agrupamento, as informações individuais são perdidas, restando apenas aquelas referentes à média dos grupos. Porém, quando se trabalha com muitos indivíduos, o número de estimativas de similaridade/dissimilaridade obtidas é relativamente grande, o que dificulta o reconhecimento de grupos homogêneos. A opção recai sobre o agrupamento dos indivíduos com base nas dispersões, em relação a eixos cartesianos obtidos por componentes, por coordenadas principais ou variáveis canônicas. A representação no espaço por um menor número de dimensões mostra-se útil, por projetar em um gráfico uma configuração de itens, num espaço de baixa dimensão, pois isto facilita a identificação dos grupos de genótipos relacionados (Cruz, 1990; Dias, 1998). Diversos trabalhos com coordenadas principais mostraram quanto de variação entre os genótipos é revelada por meio de dados RAPD (Colombo et al., 2000; Chowdhury et al., 2002; Sera et al., 2003).

Como o emprego de técnicas multivariadas no reconhecimento da diversidade genética impõe certo grau de estrutura nos dados, é importante que diferentes critérios de agrupamento sejam utilizados, e que se considere como correta a estrutura resultante da maior parte deles, para se assegurar que o resultado obtido não seja um artefato da técnica utilizada.

O objetivo deste trabalho foi comparar diferentes técnicas de agrupamento (hierárquica e de otimização) e análise de coordenadas principais, no estudo da divergência genética de genótipos de gergelim, mediante uso de marcadores RAPD.

## Material e Métodos

Foram utilizados 35 acessos de gergelim, pertencentes ao Banco Ativo de Germoplasma da Embrapa Algodão (Tabela 1), os quais foram cultivados em área experi-

mental do Departamento de Produção Vegetal da Unesp, Campus de Jaboticabal, SP.

Vinte dias depois da semeadura, duas folhas jovens foram coletadas em cada dez plantas, por genótipo. O processamento do tecido foliar e a extração de DNA para o marcador RAPD foram efetuados pelo protocolo CTAB, com modificações (Arriel et al., 2002). No processo de otimização das reações de amplificação de RAPD foram testados 76 iniciadores em sete cultivares de gergelim, e aqueles que mostraram reprodutibilidade de bandas, num total de cinco repetições por iniciador, foram selecionados, tendo-se adotado os critérios de repetibilidade, na qualidade e quantidade de bandas amplificadas.

O volume da reação de PCR foi de 25  $\mu$ L, com tampão Tris-HCL 10 mM, pH 8 e 50 mM KCl, 2,5 mM

MgCl<sub>2</sub>, 1 unidade de *Taq* DNA Polimerase, 0,2 mM de dNTPs, 30 ng do iniciador, 40 ng de DNA genômico e água Milli-Q. As condições de amplificação envolveram uma etapa inicial de desnaturação a 94°C por três minutos, seguida de 42 ciclos de desnaturação a 94°C por um minuto; anelamento do iniciador a 37°C por um minuto; e extensão a 72°C por três minutos. Ao término das reações, as amostras foram mantidas a 4°C.

Os fragmentos de RAPD foram fracionados em gel de agarose 1,5% em tampão TBE 1X (Tris base 1 M, ácido bórico e EDTA 500 mM), corados com 3  $\mu$ L de brometo de etídio (10 mg mL<sup>-1</sup>) e, em seguida, visualizados sob luz ultravioleta. Os tamanhos dos fragmentos amplificados foram estimados, por comparação, com marcador de peso molecular 100 pb, DNA Ladder.

Os 55 iniciadores da Operon Technologies (Tabela 2) foram utilizados para gerar uma matriz binária, a partir da codificação 1 para presença e 0 para ausência da banda. As estimativas de similaridade entre todos os pares de acessos (ii') foram convertidas em dissimilaridade por meio do complemento aritmético (D) do coeficiente de similaridade de Jaccard (Cruz & Carneiro, 2003):  $S_{ii'}$  Jaccard =  $a/(a + b + c)$ , em que a é o número de concordâncias do tipo 1 1; b é o número de discordâncias do tipo 1 0; e c é o número de discordâncias do tipo 0 1, sendo  $D_{ii'} = 1 - S_{ii'}$ .

Esse coeficiente foi escolhido por suas propriedades matemáticas, que desconsideram a ausência de bandas como sinônimo de similaridade genética e atribuem diferentes pesos à presença conjunta de bandas (Dias, 1998; Cruz & Carneiro, 2003).

O programa NTSYS-pc (Rohlf, 1992) foi empregado para a aplicação das técnicas de aglomeração hierárquica do vizinho mais próximo, vizinho mais distante da média aritmética entre pares não ponderados (UPGMA) e da análise das coordenadas principais. Foram usados o método aglomerativo de otimização de Tocher (Cruz, 2001) e o das coordenadas principais (Rohlf, 1992).

Nos métodos de agrupamentos hierárquicos, os genótipos foram agrupados de forma sucessiva, com base em suas distâncias mínimas, máximas ou pela média aritmética não ponderada entre os pares de genótipos, em que para cada grupo formado foi necessário reconstruir uma nova matriz de dissimilaridade, até a formação de um único grupo. Os passos dos agrupamentos foram representados na forma de dendrogramas e a comparação dos métodos de agrupamento foi realizada a partir do coeficiente de correlação cofenética (CCC), que traduz a relação momento-produto, calculado entre

**Tabela 1.** Identificação dos genótipos de gergelim, com suas respectivas denominações e procedências.

Identificação	Denominação	Procedência
Genótipo 1	Guatemala	Guatemala
Genótipo 2	Seridó 1	Brasil
Genótipo 3	Nicarágua	Nicarágua
Genótipo 4	Venezuela	Venezuela
Genótipo 5	Paquistão	Paquistão
Genótipo 6	Mexicana	México
Genótipo 7	CNPA G2	Brasil
Genótipo 8	CNPA G3	Brasil
Genótipo 9	CNPA G4	Brasil
Genótipo 10	FAO 52593	Estados Unidos
Genótipo 11	Arawaca 2	Venezuela
Genótipo 12	Arawaca 4	Venezuela
Genótipo 13	IAPAR 320	Venezuela
Genótipo 14	IAPAR 322	Venezuela
Genótipo 15	Valls et al. 7834	Brasil
Genótipo 16	Valls et al. 7835	Brasil
Genótipo 17	SB IMPROVED BACO	Estados Unidos
Genótipo 18	SB-S-9-LP-85	Estados Unidos
Genótipo 19	TMV5	Estados Unidos
Genótipo 20	TMV6	Estados Unidos
Genótipo 21	Amostra 3 Campina Grande/181	Brasil
Genótipo 22	Amostra 3 Campina Grande/182	Brasil
Genótipo 23	Amostra 1 Campina Grande/183	Brasil
Genótipo 24	Picos 05	Brasil
Genótipo 25	Fazenda Viola	Brasil
Genótipo 26	Amostra 3/203	Brasil
Genótipo 27	Amostra 05/204	Brasil
Genótipo 28	Currais Novos 06	Brasil
Genótipo 29	Jericó 1	Brasil
Genótipo 30	VCR-101	Estados Unidos
Genótipo 31	GP 3314	Brasil
Genótipo 32	Haste roxa ramificado	Estados Unidos
Genótipo 33	Variedade indeiscente 1	Venezuela
Genótipo 34	Variedade indeiscente 2	Venezuela
Genótipo 35	Variedade indeiscente 3	Venezuela

os elementos da matriz original e os da matriz cofenética, resultante da simplificação proporcionada pelo método

**Tabela 2.** Relação de iniciadores selecionados, com respectivos nomes comerciais, seqüência de bases nitrogenadas, número total de fragmentos (NTF), número de fragmentos polimórficos (NFP) e porcentagem de polimorfismo (P%) obtidos nas reações RAPD, em 35 genótipos de gergelim.

Iniciador	Seqüência	NTF	NFP	P%
OPM-01	5'-GTTGGTGGCT-3'	8	2	25,00
OPM-02	5'-ACAACGCCTC-3'	15	10	66,66
OPM-04	5'-GGCGTTGTC-3'	14	7	50,00
OPM-06	5'-CTGGGCAACT-3'	21	9	42,86
OPM-07	5'-CCGTGACTCA-3'	20	6	30,00
OPM-09	5'-GTCTTGCGGA-3'	15	9	60,00
OPM-10	5'-TCTGGCGCAC-3'	12	6	50,00
OPM-11	5'-GTCCACTGTG-3'	15	9	60,00
OPM-12	5'-GGGACGTTGG-3'	15	8	53,33
OPM-13	5'-GGTGGTCAAG-3'	16	6	37,50
OPM-14	5'-AGGGTCGTTT-3'	18	11	61,11
OPM-15	5'-GACCTACCAC-3'	18	10	55,55
OPM-16	5'-GTAACCAGCC-3'	14	8	57,14
OPM-17	5'-TCAGTCCGGG-3'	15	7	46,66
OPM-18	5'-CACCATCCGT-3'	13	8	61,53
OPM-19	5'-TCTGACGGG-3'	15	7	46,66
OPN-02	5'-ACCAGGGGCA-3'	15	10	66,66
OPN-03	5'-GGTACTCCCC-3'	16	12	75,00
OPN-04	5'-GACCGACCCA-3'	17	7	41,17
OPN-05	5'-TCTGACGGG-3'	15	10	66,66
OPN-06	5'-GAGACGCACA-3'	13	7	53,84
OPN-07	5'-CAGCCACAGAG-3'	14	7	50,00
OPN-09	5'-TGCCGGCTTG-3'	15	7	46,66
OPN-10	5'-ACAACCTGGGG-3'	16	9	56,25
OPN-11	5'-TCGCCGCAAA-3'	15	9	60,00
OPN-12	5'-CACAGACACC-3'	16	8	50,00
OPN-13	5'-AGCGTCACTC-3'	14	7	50,00
OPN-14	5'-TCTGACGGG-3'	10	6	60,00
OPN-15	5'-CAGCGACTGT-3'	17	8	47,06
OPN-16	5'-AAGCGACCTG-3'	16	7	43,75
OPN-17	5'-CATTGGGGAG-3'	15	6	40,00
OPN-18	5'-GGTGGTCCG-3'	19	10	52,63
OPN-19	5'-GTCCGTACTG-3'	14	5	35,71
OPN-20	5'-GGTGTCCGT-3'	17	7	41,17
OPP-01	5'-GTAGCACTCC-3'	10	3	30,00
OPP-02	5'-TCGCACAGCA-3'	10	6	60,00
OPP-03	5'-CTGATACGCC-3'	9	4	44,44
OPP-04	5'-GTGTCTCAGG-3'	12	5	41,66
OPP-06	5'-GTGGGCTGAC-3'	17	7	41,17
OPP-07	5'-GTCCATGCCA-3'	13	5	38,46
OPP-08	5'-ACATCGCCCA-3'	18	9	50,00
OPP-09	5'-GTGTCCGCA-3'	16	8	50,00
OPP-12	5'-AAGGGCGAGT-3'	15	5	33,33
OPP-13	5'-GGAGTGCCTC-3'	13	8	61,53
OPP-14	5'-CCAGCCGAAC-3'	13	4	30,77
OPP-15	5'-GGAAGCCAAC-3'	15	7	46,66
OPP-16	5'-CCAAGTGCC-3'	14	4	28,57
OPP-19	5'-GGGAAGGACA-3'	14	7	50,00
OP-Q06	5'-GAGCGCCTTG-3'	7	3	42,86
OP-Q11	5'-TCTCCGCAAC-3'	9	6	66,66
OP-AW11	5'-CTGCCACAG-3'	9	4	44,44
OP-AW15	5'-CCAGTCCCAA-3'	11	5	45,45
OP-AW18	5'-GGCGCAACTG-3'	10	3	30,00
OP-AD12	5'-AAGAGGGCGT-3'	8	3	37,50
OP-AD16	5'-AACGGGCGTC-3'	8	3	37,50
Total		769	374	
Média		13,98	6,80	48,21
Porcentagem			48,64	

de agrupamento, depois da construção do dendrograma. Valores de CCC superiores a 0,8 indicam boa representatividade entre as distâncias (Bussab et al., 1990).

## Resultados e Discussão

Dos 76 iniciadores utilizados, 55 foram selecionados (Tabela 2) por apresentarem quantidade, qualidade e reprodutibilidade de bandas amplificadas. Um total de 769 bandas foi gerado, das quais 48,64% (374) foram polimórficas. O número de bandas produzidas, por iniciador, variou de 7 (OPQ06) a 21 (OPM06), com média de 14 bandas por iniciador. A taxa de polimorfismo por iniciador foi de aproximadamente sete bandas, com uma variação de 2 (OPM-01) a 12 bandas polimórficas (OPN-03). O tamanho dos fragmentos amplificados variou de 300 a aproximadamente 2.000 pb. Ressalta-se que as diferenças entre os genótipos foram determinadas em função da frequência de bandas e não da presença ou ausência de bandas específicas. Porém, deve-se considerar que algumas bandas específicas observadas em certos acessos, a partir de determinado iniciador, podem ser úteis, futuramente, na caracterização de cultivares.

Uma análise estatística descritiva das matrizes de distâncias geradas pelo coeficiente de dissimilaridade mostrou que, em média, a distância entre os pares de genótipos foi de 0,28. Destacaram-se como os pares mais divergentes os genótipos 2 (Seridó 1) e 12 (Arawaca 4), e os mais similares os genótipos 30 (VCR-101) e 31 (GP 3314). O coeficiente de variação médio foi de 10,6%, que indica uma boa precisão das estimativas. As medidas de distâncias genéticas, obtidas para a formação do agrupamento, variaram de 0,10 a 0,46, sugerindo que o germoplasma utilizado apresenta uma estreita base genética, diferentemente da variação da similaridade genética observada por Bhat et al. (1999), entre 0,19 e 0,89, com marcadores RAPD em 58 acessos de gergelim indiano e exótico, a partir de medidas de distâncias obtidas por meio do mesmo coeficiente. É importante ressaltar que o alto grau de diversidade genética, aliado às informações de caracterização morfológica, possibilitou ratificar a hipótese de que, provavelmente, a região indiana é o local de origem do gergelim cultivado.

O resultado da hierarquização dos genótipos pelos métodos do UPGMA, vizinho mais próximo e vizinho



mais distante (Figura 1) mostra as diferenças no posto dos genótipos. Nos agrupamentos hierárquicos, a delimitação dos grupos é feita de maneira subjetiva, observando-se pontos de alta mudança de nível no dendrograma e, por isso, diferentes padrões de agrupamento podem ocorrer.

Quando não se tem informação sobre a relação genética entre a maioria dos genótipos, não se pode determinar que método de agrupamento é mais acurado. Assim, ao se comparar os resultados de diferentes métodos, pode-se evitar inferências errôneas. Verifica-se que, apesar da estrutura final mostrar-se parecida, principalmente entre os métodos UPGMA e vizinho mais próximo, ocorreram alterações nos níveis em que os genótipos são agrupados.

Ao se adotar um percentual de divergência genética em torno de 36% (Figura 1), constata-se a formação de diferentes agrupamentos heteróticos dos genótipos descritos na Tabela 1. Para o método UPGMA, foram formados 11 grupos, dos quais nove são unitários, constituídos pelos genótipos 2, 12, 7, 9, 8, 6, 10, 5 e 1; um décimo grupo foi formado pelos acessos 3 e 4, e um grande grupo formado pelos demais genótipos.

No método do vizinho mais próximo, podem ser observados quatro grupos, sendo que em três deles há grupos únicos dos genótipos 2, 7 e 9, e o quarto grupo foi constituído pelos demais. Para o método do vizinho mais distante, 15 grupos foram formados, dos quais 11 se constituíram de grupos unitários dos genótipos 2, 12, 9, 8, 7, 13, 21, 10, 6, 5 e 1, um grupo de dois acessos, constituído pelos genótipos 3 e 4, outro grupo formado pelos genótipos 35, 34, 33 e 32, um grupo com os genótipos 22, 20, 17, 25, 19, 18 e 11 e o último formado pelos genótipos 31, 30, 29, 26, 24, 23, 16, 28, 15, 27 e 14.

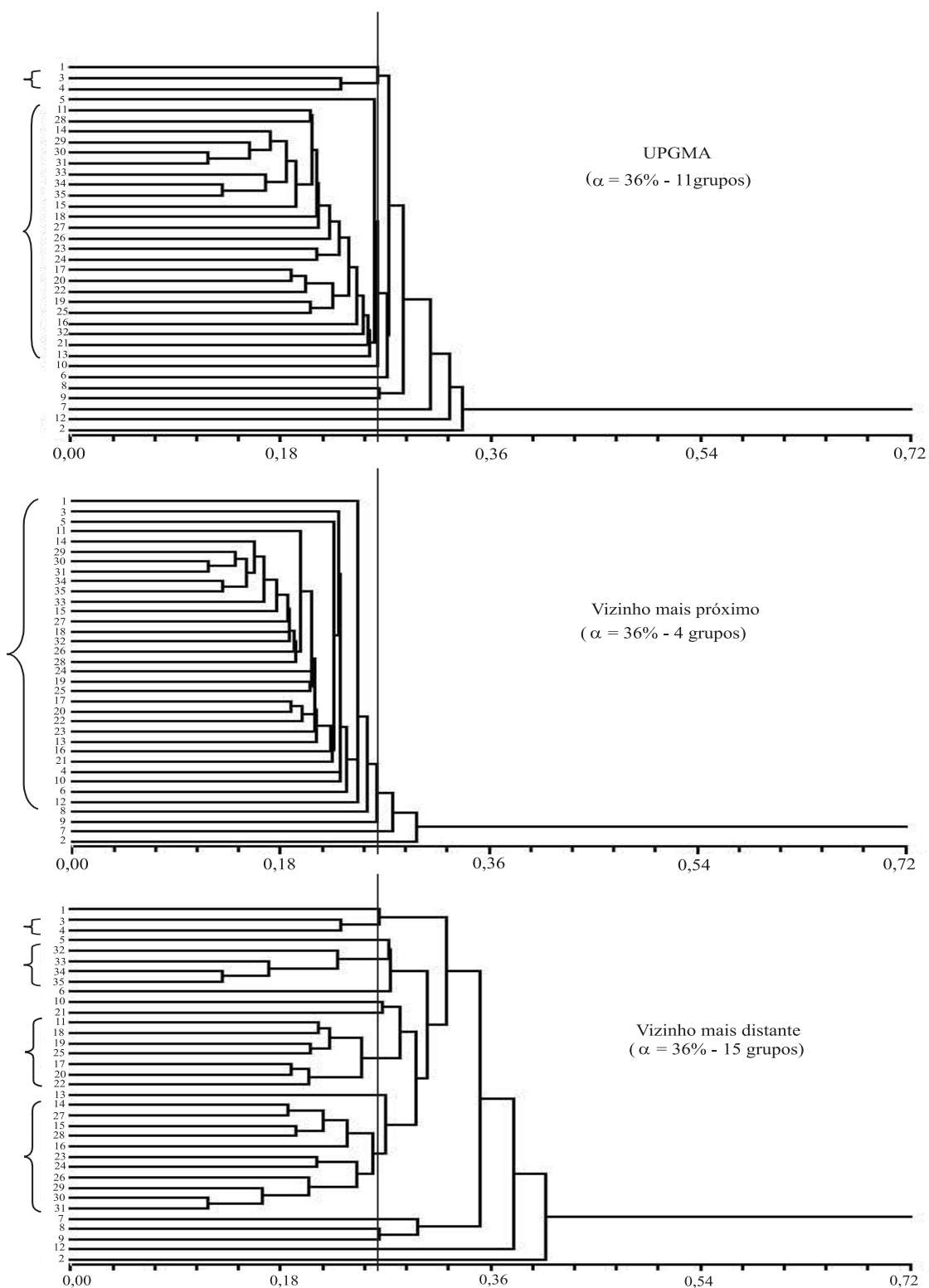
Entre os genótipos avaliados, os que se posicionam nos pontos extremos do dendrograma, com exceção do genótipo 12, são as cultivares de gergelim mais utilizadas comercialmente; 1, 3, 4, 5 e 6 são variedades originadas de outros países; e os genótipos 2, 7, 8, e 9 são cultivares nacionais, lançadas pela Embrapa Algodão. Isto sugere que, apesar da modesta diversidade imposta aos materiais, pela pressão da seleção adotada nos programas de melhoramento de cada país, no âmbito molecular esses materiais apresentam-se relativamente dissimilares. Note-se que, em todos os procedimentos de agrupamento utilizados, o genótipo 2, representado pela cultivar Seridó 1, formou um grupo único, em rela-

ção aos outros genótipos. Entre todos os genótipos, esta cultivar se diferenciou por apresentar o menor número de bandas.

Para se avaliar o grau de ajuste entre as matrizes de dissimilaridade e as matrizes resultantes dos agrupamentos, para formação dos dendrogramas dos três métodos, compararam-se as estimativas dos coeficientes de correlação cofenética (CCC). Quanto maior o valor do coeficiente, menor é a distorção provocada ao se agrupar os genótipos (Bussab et al., 1990). O método UPGMA proporcionou excelente representação das distâncias na forma de dendrogramas (CCC = 0,89), seguido pelo método do vizinho mais próximo (CCC = 0,85) e pelo método do vizinho mais distante, que apresentou CCC = 0,79. A superioridade do UPGMA já foi relatada por Dias (1998).

No agrupamento pelo método de otimização de Tocher, ocorreu a formação de quatro grupos, três deles constituídos por apenas um genótipo (2, 12 ou 7). O genótipo 2 representa a cultivar Seridó 1, de hábito de crescimento ramificado, ciclo tardio, uma cápsula por axila foliar, sementes de coloração creme e suscetibilidade às principais doenças da cultura; o genótipo 7 é a CNPA G2, que tem por características a presença de três frutos por axila foliar, ciclo médio e tolerância à mancha angular. O genótipo 12 representa a cultivar Arawaca 4, de porte médio, cápsulas pequenas e sementes de coloração marrom e apresentou tolerância às infestações de doenças e praga.

Os demais genótipos formaram o quarto grupo, que compreendeu aproximadamente 91% do germoplasma avaliado, confirmando a baixa amplitude de variabilidade genética. No método Tocher, um indivíduo ainda não agrupado só é incluído em determinado grupo em formação se sua distância média em relação a esse grupo não ultrapassar determinado valor preestabelecido. Tal valor geralmente é tomado como a maior amplitude do conjunto das menores estimativas de distância que envolvem cada um dos indivíduos em agrupamento. Por isso, a grande divergência dos genótipos 2, 7 e 12 confirma sua separação em grupos distintos. Um ponto a ser considerado é a relação da procedência geográfica como indicador de diversidade genética, pois, em muitos casos, a separação geográfica geralmente não resulta em maior distância genética (Cruz & Carneiro, 2003), o que foi constatado, neste trabalho, pela presença de genótipos de diferentes países que compuseram o grupo quatro, onde se encontra a maioria do germoplasma.



**Figura 1.** Dendrogramas dos diferentes métodos de agrupamentos construídos a partir do complemento aritmético do coeficiente de Jaccard entre 35 genótipos de gergelim, considerando os marcadores moleculares RAPD. A numeração dos genótipos está de acordo com a Tabela 1.

Ao comparar os grupos formados pelo método de otimização de Tocher com os dendrogramas, verifica-se que o método UPGMA é o que mais se assemelha na diferenciação dos genótipos e, dependendo do ponto de corte escolhido, podem ser formados basicamente quatro grupos, enquanto que nos demais, constata-se diferenças na hierarquização inicial dos genótipos.

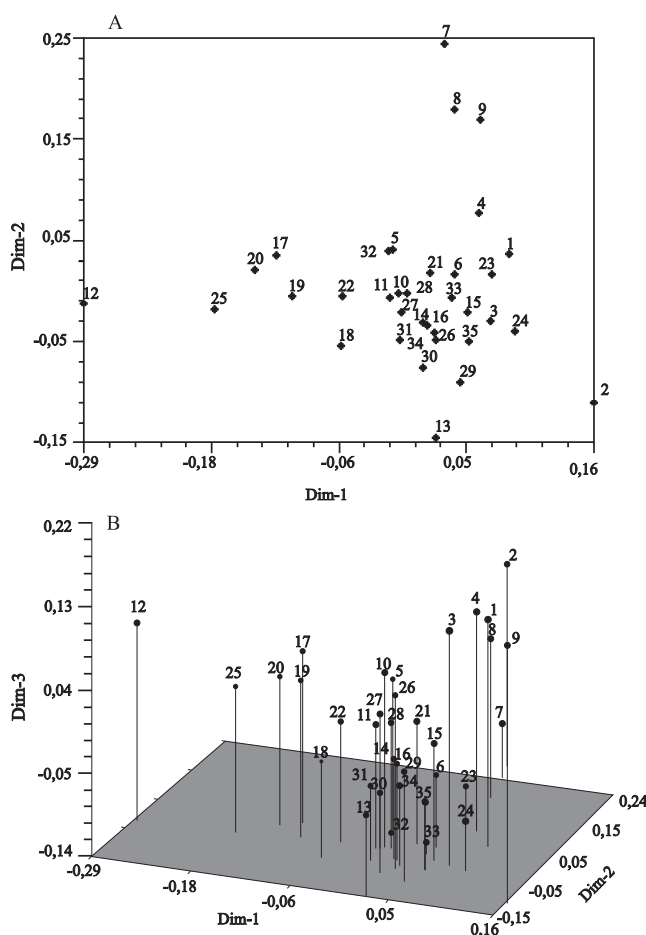
Para averiguar se o agrupamento de genótipos apresentaria melhor resolução, empregou-se a técnica de coordenada principal (PCA), para análises de dados RAPD. Por meio das análises, constatou-se que a porcentagem da variabilidade acumulada nos primeiros eixos (autovalores) gerados não é grande, o que ratifica a pequena diversidade genética do germoplasma em estudo. Mesmo com esta limitação, foram gerados gráficos bidimensionais e tridimensionais, que relacionam os novos escores para os genótipos, em função dos dois primeiros eixos definidos na análise (Figura 2).

A representação gráfica da associação dos genótipos, a partir da projeção num plano espacial, procura reduzir a dimensionalidade dos espaços interpontos. Entretanto, a PCA não persegue a maximização da variação, nas projeções dos pontos no espaço considerado, e requer apenas uma relação monotônica entre as distâncias originais e distâncias interpontos. Em razão disso, os primeiros três componentes da análise das coordenadas principais explicaram 35,13% de variação entre as amostras, com 18,59, 8,64 e 7,90% para o primeiro, segundo e terceiro componentes, respectivamente. Para explicação de mais de 81% de variação entre os genótipos, foram necessários 18 autovalores. A contribuição dos componentes restantes, individualmente, foi menos de 1,5% cada.

Na avaliação de 58 acessos de gergelim, Bhat et al. (1999) observaram que os dois primeiros componentes principais responderam por 21% da variação total, e para explicar 75% da variabilidade, seriam necessários 18 autovalores. Por sua vez, Colombo et al. (2000), ao avaliar 126 genótipos de mandioca, verificaram que os dois primeiros eixos das coordenadas principais foram responsáveis por apenas 14% da variação existente, porém foi possível separar os genótipos em quatro grupos. Chowdhury et al. (2002) constataram que, em 48 cultivares de soja, os dois primeiros componentes principais foram responsáveis por 45,95% da variação total. Sera et al. (2003) avaliaram 14 genótipos elites de café e observaram que as três primeiras coordenadas principais foram responsáveis por 53,8% da variação dos dados. De acordo com os resultados pode-se

concluir que, independentemente da variabilidade encontrada, a projeção dos pontos nos espaços bidimensional e tridimensional permitiu representar toda a variabilidade existente.

A projeção das coordenadas principais (Figura 2), além de reforçar o posicionamento dos genótipos mais distintos (2, 7 e 12), revelou que o genótipo 13 também é divergente dos demais. É possível, também, identificar um grupo isolado constituído dos acessos 25, 20, 19 e 17, e dois grupos independentes, formados pelos genótipos 18 e 22, 8 e 9. O genótipo 4, nos métodos de hierarquização do vizinho mais distante e UPGMA, revelava-se muito próximo ao 3, enquanto no método do vizinho mais próximo, comportou-se como mais independente, conforme é evidenciado na projeção no plano.



**Figura 2.** Representação bidimensional (A) e tridimensional (B) da análise de coordenadas principais dos 35 genótipos de gergelim, baseada na distância genética calculada com dados de marcadores RAPD.

Por meio da utilização dos diferentes critérios de agrupamento, pode-se identificar a existência de agrupamento natural dos genótipos, em função da similaridade genética do germoplasma avaliado. Entretanto, o emprego de mais de um método de agrupamento, em razão das diferenças na hierarquização, otimização e ordenação dos grupos, permite que a classificação deles se complemente em função dos critérios que cada técnica utiliza, e impede que inferências errôneas sejam adotadas na alocação de materiais, dentro de um determinado subgrupo de genótipos. Pode-se concluir que os métodos de hierarquização UPGMA e otimização de Tocher devem ser empregados em conjunto, para complementar os resultados e, caso haja um número elevado de genótipos, a dispersão gráfica dos escores originados pelas coordenadas principais auxiliará numa melhor distinção dos agrupamentos formados.

A baixa diversidade entre esses genótipos pode ser atribuída ao fato de o gergelim ser uma cultura introduzida, diferentemente da coleção de gergelim do continente indiano, centro de domesticação da espécie, onde Bhat et al. (1999) estimaram uma maior diversidade. Essa observação pode ser reforçada pelos resultados obtidos por Isshiki & Umezaki (1997) em que, dos 11 marcadores bioquímicos usados para avaliação de 41 acessos de gergelim cultivado, originários do Japão, Coréia e Tailândia, a variação foi reportada somente em um único sistema enzimático.

De modo geral, os resultados dos métodos de agrupamentos e da análise de coordenadas principais evidenciaram que os marcadores RAPD foram eficientes na caracterização da diversidade genética entre os genótipos de gergelim, e que, posteriormente, poderão ser utilizados na caracterização da coleção de germoplasma disponível na Embrapa Algodão, com vistas à simplificação do banco de germoplasma e formação de "Core Collection", uma vez que, pelos resultados obtidos, apesar do pequeno número de acessos avaliados, percebe-se a existência de duplicatas na coleção.

Os grupos únicos, representados pelas cultivares comerciais, poderiam ser usados como teste para capacidade de combinação, pois genótipos divergentes podem ser potenciais genitores, em termos de valor genético (efeito médio dos seus genes) a ser passado para os seus descendentes, possibilitando que novas combinações alélicas possam ser obtidas e selecionadas, mediante retrocruzamentos e seleções recorrentes.

Outro aspecto a ser abordado é o polimorfismo encontrado em alguns iniciadores como OPM09, OPN03, OPN18 e OPP08, que poderão ser úteis na averiguação de fragmentos moleculares específicos por genótipo, após sua conversão em regiões de seqüências específicas (SCARs).

É importante ampliar a base genética do gergelim, a fim de se explorar a máxima variabilidade na seleção de populações segregantes que pode ser alcançada pelo uso de genótipos mais divergentes. O conhecimento da relação genética será de grande valor nas diretrizes de exploração do gergelim, pois possibilitará a otimização do uso do germoplasma disponível.

### Conclusões

1. As cultivares Seridó 1 e Arawaca 4 são as mais divergentes geneticamente, e VCR-101 e GP-3314 são as mais similares.
2. A hierarquização dos genótipos é alterada em função dos diferentes métodos usados.
3. Entre os métodos de agrupamento hierárquicos, o UPGMA apresenta o melhor ajuste para as distâncias originais e estimadas.
4. As análises das coordenadas principais confirmam a baixa diversidade entre os genótipos avaliados.
5. O método UPGMA, o método de otimização de Tocher e as análises de coordenadas principais são complementares na formação dos agrupamentos.

### Referências

- ARRIEL, N.H.C.; FREIRE, E.C.; ANDRADE, F.P. Melhoramento genético. In: BELTRÃO, N.E. de M.; VIEIRA, D.J. (Ed.). **O agronegócio do gergelim no Brasil**. Brasília: Embrapa Informação Tecnológica; Campina Grande: Embrapa Algodão, 2001. p.247-284.
- ARRIEL, N.H.C.; UNÊDA-TREVISOLI, S.H.; CAPELATO, A.; DI MAURO, S.M.Z.; DI MAURO, A.O. Análise comparativa de quatro protocolos de extração de DNA genômico em *Sesamum indicum* L. **Revista Brasileira de Oleaginosas e Fibrosas**, v.6, p.525-535, 2002.
- BERED, F.; BARBOSA NETO, J.F.; ROCHA, B.M. da; CARVALHO, F.I.F. de. Genetic variability in wheat (*Triticum aestivum* L.) germplasm revealed by RAPD markers. **Crop Breeding and Applied Biotechnology**, v.2, p.495-502, 2002.
- BHAT, K.V.; BABREKAR, P.P.; LAKHANPAUL, S. Study of genetic diversity in Indian and exotic sesame (*Sesamum indicum* L.)



- germplasm using random amplified polymorphic DNA (RAPD) markers. **Euphytica**, v.110, p.21-34, 1999.
- BUSSAB, W. de O.; MIAZAKI, E.S.; ANDRADE, D.F. **Introdução à análise de agrupamentos**. São Paulo: Associação Brasileira de Estatística, 1990. 105p.
- CHOWDHURY, A.K.; SRINIVES, P.; TONGPAMNAK, P.; SAKSOONG, P.; CHATWACHIRAWONG, P. Genetic relationship among exotic soybean introductions in Thailand: consequence for varietal registration. **Science Asia**, v.28, p.227-239, 2002.
- COLOMBO, C.; GÉRARD, S.; CHARRIER, A. Diversity within American cassava germplasm based on RAPD markers. **Genetics and Molecular Biology**, v.23, p.189-199, 2000.
- COSTA, M.R.; CARDOSO, E.R.; OHAZE, M.M.M. Similaridade genética de cultivares de mandioca (*Manihot esculenta*) por meio de marcadores RAPD. **Ciência e Agrotecnologia**, v.27, p.158-164, 2003.
- CRUZ, C.D. **Aplicação de algumas técnicas multivariadas no melhoramento de plantas**. 1990. 188p. Tese (Doutorado) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- CRUZ, C.D. **Programa Genes**: aplicativo computacional em genética e estatística. Versão 2003. Viçosa: UFV, 2001. 648p.
- CRUZ, C.D.; CARNEIRO P.C.S. **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, 2003. v.2. 585p.
- DIAS, L.A. dos S. Análises multidimensionais. In: ALFENAS, A.C. (Ed.). **Eletróforese de isoenzimas e proteínas afins**: fundamentos e aplicações em plantas e microrganismos. Viçosa: UFV, 1998. p.405-475.
- EMYGDIO, B.M.; ANTUNES, I.F.; NEDEL, J.L.; CHOER, E. Diversidade genética em cultivares locais e comerciais de feijão baseada em marcadores RAPD. **Pesquisa Agropecuária Brasileira**, v.38, p.1165-1171, 2003.
- ISSHIKI, S.; UMEZAKI, T. Genetic variations of isozymes in cultivated sesame (*Sesamum indicum* L.). **Euphytica**, v.93, p.375-377, 1997.
- MANIVANNAN, N.; NADARAJAN, N. Genetic divergence in sesame. **Madras Agricultural Journal**, v.83, p.789-790, 1996.
- OLIVEIRA, R.P. de; CRISTOFANI, M.; AGUILAR-VILDOSO, C.I.; MACHADO, M.A. Diversidade genética entre híbridos de tangerina 'Cravo' e laranja 'Pera'. **Pesquisa Agropecuária Brasileira**, v.37, p.479-484, 2002.
- PEREIRA, C.D.; KERR, W.E. Divergência genética entre doze genótipos de abacaxizeiro (*Ananas comosus* L. Merrill.) estimada por análise de marcadores RAPD. **Revista Brasileira de Fruticultura**, v.23, p.335-338, 2001.
- ROHLF, F.J. **NTSYS-pc**: numerical taxonomy and multivariate analysis system: version 1.10. New York: Exeter Software, 1992. 470p.
- SANZ-CORTÉS, F.; BADENES, M.L.; PAZ, S.; ÍÑIGUEZ, A.; LLÁCER, G. Molecular characterization of olive cultivars using RAPD markers. **Journal of the American Society for Horticultural Science**, v.126, p.7-12, 2001.
- SERA, T.; RUAS, P.M.; RUAS, C. de F.; DINIZ, L.E.C.; CARVALHO, V. de P.; RAMPIM, L.; RUAS, E.A.; SILVEIRA, S.R. da. Genetic polymorphism among 14 elite *Coffea arabica* L. cultivars using RAPD markers associated with restriction digestion. **Genetics and Molecular Biology**, v.26, p.59-64, 2003.
- SINGH, R.M.; SINGH, A.K.; ARKASH, K.; THAKRAL, N.K.; KUMAR, P. Genetic divergence in sesame. **Annals of Biology Ludhiana**, v.13, p.41-45, 1997.
- SWAIN, D.; DIKSHIT, U.N. Genetic divergence in rabi sesame (*Sesamum indicum* L.). **Indian Journal of Genetics and Plant Breeding**, v.57, p.296-300, 1997.

---

Recebido em 23 de setembro de 2004 e aprovado em 10 de janeiro de 2006