

# COMPARAÇÃO DE MÉTODOS DE AGRUPAMENTO PARA O ESTUDO DA DIVERSIDADE GENÉTICA DE CULTIVARES DE ARROZ<sup>1</sup>

José Jonas Pereira<sup>2</sup>  
Cosme Damião Cruz<sup>3</sup>

## RESUMO

Foi avaliada a diversidade genética de cultivares de arroz de várzea úmida por variáveis canônicas e análise de agrupamento. Na análise de agrupamento foram utilizados os métodos de Ward, do vizinho mais próximo, da ligação média e de Tocher, com base nas distâncias euclidianas e de Mahalanobis. O padrão de agrupamento divergiu tanto em relação aos métodos utilizados quanto às medidas de distância. A análise de agrupamento, com base na distância de Mahalanobis, formou quatro grupos com todos os métodos. Com base na distância euclidiana, o número de grupos variou entre seis e 10, segundo o método utilizado. Apesar da diferença do número de grupos e do padrão de agrupamento, de maneira geral os resultados foram concordantes. Tanto o padrão de agrupamento quanto as combinações indicadas para cruzamento são mais consistentes quando os métodos são utilizados com a mesma medida de distância.

Palavras-chave: *Oryza sativa*, análise multivariada.

## ABSTRACT

### COMPARISON OF CLUSTERING METHODS FOR THE STUDY OF GENETIC DIVERSITY IN RICE CULTIVARS

Genetic diversity of humid lowland rice cultivars were estimated by using canonic variables and cluster analysis. The latter included Ward methods, single linkage, average linkage and Tocher's, with Euclidean and Mahalanobis distances. The clustering patterns

---

<sup>1</sup> Aceito para publicação em 07.10.2002.

<sup>2</sup> Depto. de Estatística da Universidade Federal de Juiz de Fora. 36036-330 Juiz de Fora, MG. E-mail: [jjonas@estatistica.ufjf.br](mailto:jjonas@estatistica.ufjf.br)

<sup>3</sup> Depto. de Biologia Geral da Universidade Federal de Viçosa. 36571-000 Viçosa, MG. E-mail: [cdcruz@ufv.br](mailto:cdcruz@ufv.br)

obtained diverged both in the applied methods and distances. The Mahalanobis distance-based clustering analysis discriminated four groups regardless of the applied method. On the other hand, the Euclidean-based clustering analysis ranged from six to ten groups, depending on the method. Despite the varying number of groups and the clustering pattern, the results showed an overall agreement. The clustering patterns and the combinations indicated for crossing are more consistent when methods are based on similar types of distance.

Key words: *Oryza sativa*, multivariate analysis,

## INTRODUÇÃO

A escolha apropriada de genitores possibilita a obtenção de progênes superiores e é essencial para o sucesso de qualquer programa de hibridação convencional. Dentro de determinada faixa de diversidade, é aceito que cruzamentos de pais mais distantemente relacionados mostram maior heterose que os de pais mais estreitamente relacionados. Desse modo, para que os resultados dos programas de hibridação sejam conseguidos mais rapidamente, é necessário que na população-base para a seleção se alie alta média (se o que se deseja é alta média) para o caráter a ser selecionado e ampla variabilidade genética. O relacionamento entre os pais divergentes e a heterose para características complexas, como é o caso da produção de grãos, pode também ser visto como consequência da complementação de características componentes (8, 12).

Em geral, a análise de divergência é uma tentativa de identificar pais adequados para a obtenção de heterose e para a recombinação em programas de melhoramento. No entanto, quantificar o grau de associação entre heterose nos híbridos e divergência genética nos genitores não constitui tarefa muito simples. Sendo essa divergência, em geral, desconhecida, é necessário determinar o seu nível, empiricamente, por meio de cruzamentos.

No estudo da divergência genética, o grau de dissimilaridade entre os indivíduos dentro ou entre espécies, ou entre genótipos dentro de uma população melhorada, pode ser estimado por meio de técnicas multivariadas, como: análise de componentes principais, variáveis canônicas e análise de agrupamento (*cluster analysis*). Tais análises podem revelar grupos de genótipos similares quando plotados em gráficos de dispersão (componentes principais e variáveis canônicas). A análise de agrupamento busca, em princípio, identificar grupos de pontos no espaço. Os genótipos dentro de um mesmo grupo devem estar relacionados (10).

No Brasil, segundo Rangel e Neves (14), os cultivares de arroz irrigado mais amplamente utilizados são provenientes de apenas sete ancestrais, apresentando, portanto, estreita base genética. Essa base genética estreita é fator limitante para os programas de seleção,

diminuindo as chances para recombinação e propiciando pequeno ganho genético.

A presente investigação teve como objetivos obter informações sobre a diversidade genética de cultivares de arroz dos tipos “moderno” e “tradicional”, utilizando-se para isso diferentes medidas de dissimilaridade e técnicas de agrupamento; avaliar os fatores que influenciam a divergência genética; obter grupos de genitores mais ou menos homogêneos nos quais exista possibilidade de se fazerem escolhas apropriadas de pais que, provavelmente, produzirão segregantes agronomicamente superiores nos cruzamentos, para o desenvolvimento de variedades; e identificar cultivares similares, com vistas à otimização de conservação em bancos de germoplasma.

## MATERIAL E MÉTODOS

Foram utilizados dados de um ensaio de competição entre 49 cultivares de arroz (*Oryza sativa* L.) conduzido na EMBRAPA/CNPAP, em sistema de várzea úmida, sendo 18 desses cultivares considerados tradicionais e os demais, modernos (Quadro 1).

O delineamento utilizado foi o de blocos ao caso, com duas repetições. A área total da parcela foi de 24,0 m<sup>2</sup>, com 10 sulcos de seis metros de comprimento (4,0 x 6,0 m) por parcela e espaçamento entre sulcos de 0,40 m. A área útil da parcela foi estabelecida por seis sulcos centrais, deixando 0,50 m nas extremidades dos sulcos (2,40 x 5,0 m = 12,0 m<sup>2</sup>). A densidade de semeadura foi de 70 sementes viáveis por metro.

Foram avaliadas as seguintes características: produção de grãos em kg/ha (PROD), número de perfilhos por m<sup>2</sup> (PERF), número de panícula por m<sup>2</sup> (PAN), número de espiguetas por panícula (ESP), percentagem de fertilidade (PFERT), comprimento de panícula (CPAN), peso de 100 grãos (P100) e altura de planta (ALT). Esses dados foram obtidos em uma amostra de cinco plantas competitivas dentro da área útil da parcela.

Outras características (dados fisiológicos) foram avaliadas: índice de área foliar (IAF), peso total seco (PTS), índice de colheita (IC) e dias até floração (DF).

A diversidade genética dos cultivares foi avaliada por procedimentos multivariados: análises de variáveis canônicas e análise de agrupamento, usando o programa GENES (6). Inúmeros estudos têm mostrado a associação entre a diversidade genética dos progenitores, avaliada por essas técnicas multivariadas, e a heterose nos híbridos.

**QUADRO 1 - Cultivares de arroz dos tipos tradicionais e modernos cultivados em sistema de várzea úmida**

Cultivar	Identificação	Grupo <sup>1</sup>	Cultivar	Identificação	Grupo	Cultivar	Identificação	Grupo
T-18	Lageado	I	M-39	CR 1002	II	N-25	Cica 8	III
T-47	Come Cru	I	M-4	BR 51-282-8	II	N-21	LAC 899	III
T-2	De Abril	I	M-6	MR 15	II	N-41	BR 1	III
T-38	Desconhecido	I	M-28	IR 4829-89-2	II	N-27	Taichung Sen 3	III
T-9	C 1016-1	I	M-1	Hn 351-1-8-6	II	N-12	BG 90-2	III
T-5	Cuchilão	I	M-16	B 2039-C-Kn	II	N-29	IET 5389	III
T-10	Marabá	I	M-43	BR 51-46-5	II	N-7	IR 28-23-103-5-1	III
T-33	Skrivimankote	I	M-8	C 1117-2	II	N-30	IR 2307-247	III
T-14	Paga Dívida	I	M-20	C 168	II	N-32	Cica 9	III
T-45	Barriga Branca	I	M-26	IR 7141-18-3	II	N-42	IR 2070-423	III
T-22	Coqueiro Casca Branca	I	M-3	Mala 515	II	N-40	IET 5552	III
T-35	Douridão/Amarelão	I	M-36	C 424-2	II	N-31	IET 4247	III
T-34	Buriti	I	M-23	IR 34	II	N-13	BPI-RJ-2	III
T-15	Amarelão	I	M-37	CNM 20	II	N-46	BR IRGA 409	III
T-11	Quebra Cacho	I	M-44	IR 4859-38	II	N-49	IR 9129-2-2	III
T-24	Agulha	I				N-17	Diwani	III
T-19	Bregeiro/Nenezinho	I						
T-48	Iguape Sem Arista	I						

<sup>1</sup>/ Grupos: I cultivares tradicionais; II e III cultivares modernos

*Diversidade genética baseada em análise de variáveis canônicas*

A análise de variáveis canônicas, à semelhança de componentes principais, visa simplificar estruturalmente o conjunto de dados, isto é, reduzir a dimensionalidade dos dados, de modo que a diversidade, influenciada a princípio por um conjunto multidimensional, possa ser avaliada por um complexo bi ou tridimensional de mais fácil interpretação geométrica. Assim, a análise busca condensar um complexo de múltiplas variáveis em uma ou poucas variáveis canônicas, que são combinações lineares das variáveis originais, sendo determinadas de modo que a variação entre grupos seja maximizada em relação à variação dentro de grupos, sujeitas à restrição de serem não-correlacionadas entre si.

*Diversidade genética baseada em análise de agrupamento*

A análise de agrupamento procura discriminar geneticamente os indivíduos, ou objetos, permitindo separá-los por grupos, pela análise de um conjunto de características inerentes a cada indivíduo, ou seja, ela tem por fim agrupar, por algum critério de classificação, indivíduos, ou qualquer objeto, de forma que exista homogeneidade dentro do grupo e heterogeneidade entre grupos (13, 19). Na análise de agrupamento foram utilizados os métodos de Ward, do vizinho mais próximo, da ligação média e de Tocher, com base nas distâncias euclidianas e de Mahalanobis.

**RESULTADOS E DISCUSSÃO***Comparação do desempenho dos cultivares*

A análise de variância dos dados revelou suficiente variabilidade genética no material nas características consideradas, revelando também, em geral, ampla variação entre grupos em quase todos os caracteres (Quadro 2). Os cultivares mais produtivos (Quadro 3) são, de modo geral, os que apresentam maior perfilhamento (PERF), maior número de panícula por m<sup>2</sup> (PAN), maior número de espiguetas por panícula (ESP), menor peso de grãos (P100), boa percentagem de grãos férteis (PFERT), maior número de dias até floração (DF), alto índice de colheita (IC) e pequena altura (ALT). Dentre os cultivares mais produtivos, encontraram-se alguns que apresentavam porte elevado, como T-2, T-18, T-38, T-9 e T-47, do grupo I, e M-1 e M-4, do grupo II.

Em termos de rendimento de grãos, os cultivares dos grupos II e III, modernos, foram superiores aos tradicionais, do grupo I. Entretanto, foram observados cultivares tradicionais de boa capacidade de produção, como:

**QUADRO 2 - Resumo da análise de variância de caracteres avaliados em cultivares de arroz em várzea úmida. Grupo I - cultivares tradicionais; e grupos II e III, cultivares modernos**

		Quadrado Médio									
FV	GL	PROD <sup>(1)</sup>	PERF	PAN	ESP	PFERT	P100	ALT			
Repetições	1	30218919,14	16484,17	13730,41	12019,13	78,11	0,02	3809,46			
Tratamentos	48	1732386,38 **	12476,90 **	11547,73 **	1497,30 **	64,44 **	0,40 **	1479,74 **			
Grupos	2	10611864,00 **	173743,43 **	169814,92 **	4747,65 *	68,77 ns	5,69 **	29529,32 **			
Cultivares/Grupo I	17	1143189,47 ns	1765,54 ns	1876,96 ns	1179,53 ns	96,77 **	0,25 **	444,65 **			
Cultivares/Grupo II	14	1161939,53 ns	6080,63 **	4396,63 **	1348,21 *	38,82 *	0,14 **	267,09 **			
Cultivares/Grupo III	15	1748632,80 *	9084,07 **	8080,03 **	1563,22 **	51,14 **	0,11 **	44,72 ns			
Resíduo	48	865090,69	1908,15	1435,60	653,15	19,12	0,01	86,86			
Média		5789,73	292,30	275,35	144,36	88,07	2,76	112,86			
CV <sub>e</sub> (%)		16,06	14,94	13,76	17,70	4,96	2,59	8,26			
		Quadrado Médio									
FV	GL	CPAN	IAF	AFB	CFB	LFB	A2F	C2F			
Repetições	1	9,49	9,49	3,20	61,92	0,96	115,30	420,91			
Tratamentos	48	3,71 **	3,71 **	66,04 **	46,95 **	0,13 **	31,80 **	125,68 **			
Grupos	2	8,73 *	8,73 *	861,33 **	491,19 **	1,90 **	434,02 **	2242,08 **			
Cultivares/Grupo I	17	2,90 ns	2,90 ns	40,21 **	49,46 **	0,10 **	10,51 ns	56,84 **			
Cultivares/Grupo II	14	4,82 **	4,82 **	47,80 **	12,95 ns	0,05 **	31,52 **	18,83 ns			
Cultivares/Grupo III	15	2,91 ns	2,91 ns	6,30 ns	16,61 ns	0,01 ns	2,55 ns	21,25 ns			
Resíduo	48	1,75	1,75	8,64	10,56	0,01	6,60	19,98			
Média		22,66	22,66	7,97	24,49	1,42	8,77	36,72			
CV <sub>e</sub> (%)		5,84	5,84	36,88	13,27	8,49	29,28	12,17			

Continua...

QUADRO 2 – Continuação.

FV	GL	Quadrado Médio									
		L2F	A3F	C3F	L3F	PST	IC	DF			
Repetições	1	0,93	221,10	1035,77	0,83	20,16	225,97	78,99			
Tratamentos	48	0,14	97,69 **	209,69 **	0,12 **	4,00 **	67,80 **	133,35 **			
Grupos	2	1,98	1439,16 **	4416,00 **	1,66 **	65,41 **	802,68 **	69,67 ns			
Cultivares/Grupo I	17	0,11	39,78 **	33,17 ns	0,11 **	1,44 **	57,60 **	181,97 **			
Cultivares/Grupo II	14	0,05	75,41 **	27,23 ns	0,03 ns	1,47 *	22,27 **	96,78 **			
Cultivares/Grupo III	15	0,02	5,27 ns	19,20 ns	0,02 ns	1,07 *	23,88 **	120,87 **			
Resíduo	48	0,01	13,84	20,11	0,02	0,61	8,95	4,69			
Média		1,23	12,11	42,52	1,12	4,61	49,14	103,37			
CV <sub>e</sub> (%)		8,13	30,72	10,54	12,58	16,98	6,09	2,09			

(1) PROD - produção em kg/ha; PERF - número de perfilhos; PAN - número de panículas por m<sup>2</sup>; ESP - número de espiguetas por panículas; PFERT - percentagem de grãos cheios; P100 - peso de 100 grãos; ALT - altura da planta; CPAN - comprimento de panículas; LAF - índice de área foliar; AFB - ângulo da folha bandeira; CFB - comprimento da folha bandeira; LFB - largura da folha bandeira; A2F - ângulo da segunda folha; C2F - comprimento da segunda folha; L2F - largura da segunda folha; A3F - ângulo da terceira folha; C3F - comprimento da terceira folha; L3F - largura da terceira folha; PST - peso seco total; IC - índice de colheita; DF - dias para floração. CV<sub>e</sub> - coeficiente de variação experimental; \*, \*\* - significativos a 5% e 1%, respectivamente, pelo teste F; ns - não-significativo.

**QUADRO 3 - Resultado da análise comparativa de médias de grupos de cultivares de arroz em condições de várzea úmida. Grupo I - cultivares tradicionais; e grupos II e III - cultivares modernos**

Característica <sup>(1)</sup>	Médias		
	Grupos <sup>(2)</sup>		
	G I	G II	G III
<b>Produção</b>			
PROD	5191,90 b	6260,39 a	6021,03 a
PERF	215,28 b	323,93 a	349,28 a
PAN	199,28 b	306,17 a	332,03 a
ESP	154,51 a	146,38 ab	131,06 b
PFERT	89,61 a	86,93 a	87,42 a
P100	3,20 a	2,47 b	2,53 b
<b>Vegetativas</b>			
ALT	144,19 a	102,13 b	87,69 c
CPAN	23,21 a	22,41 a	22,28 a
DF	101,89 a	104,73 a	103,75 a
<b>Biomassa</b>			
IAF	2,92 b	3,77 a	2,95 b
PTS	6,08 a	4,15 b	3,40 c
IC	43,88 b	51,40 a	52,92 a
<b>Arquitetura</b>			
AFB	13,19 a	6,64 b	3,34 c
CFB	28,44 a	23,45 b	21,01 c
LFB	1,66 a	1,36 b	1,20 c
A2F	12,38 a	8,17 b	5,28 c
C2F	45,47 a	33,17 b	30,22 c
L2F	1,48 a	1,14 b	1,02 c
A3F	18,63 a	11,13 b	5,69 c
C3F	54,88 a	36,90 b	33,89 c
L3F	1,35 a	1,03 b	0,94 b

Médias de grupos seguidas pelas mesmas letras, na horizontal, não diferem entre si pelo teste de Scheffé, a 5% de probabilidade.

(1) PROD - produção em kg/ha; PERF - número de perfilhos; PAN - número de panículas por m<sup>2</sup>; ESP - número de espiguetas por panículas; PFERT - percentagem de grãos cheios; P100 - peso de 100 grãos; ALT - altura da planta; CPAN - comprimento de panículas; IAF - índice de área foliar; AFB - ângulo da folha bandeira; CFB - comprimento da folha bandeira; LFB - largura da folha bandeira; A2F - ângulo da segunda folha; C2F - comprimento da segunda folha; L2F - largura da segunda folha; A3F - ângulo da terceira folha; C3F - comprimento da terceira folha; L3F - largura da terceira folha; PTS - peso total seco; IC - índice de colheita; e DF - dias até a floração.

(2) G I - grupo I, com 36 cultivares; G II - grupo II, com 30 cultivares; e G III - grupo III, com 32 cultivares.



Lageado, Come Cru, De Abril, Desconhecido e C1016-1. A variabilidade genética entre os indivíduos é crucial para qualquer programa de melhoramento. Por razões de custo de manutenção, limitações de área de plantio, entre outras, genótipos são constantemente descartados, e a base genética das populações melhoradas vai se restringindo a cada estágio de um programa de melhoramento. Para recompor a variabilidade da população é preciso, muitas vezes, recorrer aos bancos de germoplasma. Na conservação do banco de germoplasma torna-se necessário identificar os genótipos similares e os divergentes, devendo manter-se apenas os últimos. Fenótipos similares podem ser produzidos por diferentes genótipos, o que torna difícil a identificação de indivíduos geneticamente similares ou não.

Num programa de melhoramento genético por hibridação, os cultivares tomados como progenitores devem ser selecionados de acordo com sua performance em relação às características desejadas e da distância genética entre eles. Espera-se que cultivares mais distantes geneticamente produzam híbridos de maior efeito heterótico.

Segundo Ghaderi et al. (8), pais geneticamente relacionados tendem a compartilhar muitos genes ou alelos em comum, e, quando dois destes pais são cruzados, há pouco estímulo, atribuído ao baixo nível de heterozigidade alélica no cruzamento. Quando dois pais são mais distantes geneticamente, assume-se que eles difiram no número de locos nos quais os efeitos de dominância estão evidentes, contribuindo para a obtenção da heterose.

#### *Diversidade genética baseada em variáveis canônicas*

Variáveis canônicas resultam de uma série de transformações ortogonais, maximizando a razão entre a variância entre grupos e a dentro de grupos. A análise de variáveis canônicas permite a visualização das diferenças entre os cultivares (ou grupos de cultivares) pela redução das dimensões do conjunto de dados, preservando a maior parte das informações biológicas. Neste estudo, a análise de variáveis canônicas foi capaz de separar claramente os tipos de cultivares tradicionais dos modernos, revelando grupos de fenótipos morfológicamente dissimilares. Os genótipos dentro de um mesmo grupo provavelmente estão relacionados. As informações extraídas da análise de variáveis canônicas podem ser utilizadas para caracterizar coleções de germoplasma e, em decisões de melhoramento, para explorar vigor híbrido ou minimizar depressão por endogamia.

Pela análise de variáveis canônicas verificou-se que as duas primeiras variáveis canônicas (VC1 e VC2) foram suficientes para explicar mais de 75% da variação. Assim, a representação da dispersão em um

gráfico bidimensional (Figura 1) pode ser utilizada numa primeira interpretação da diversidade genética entre os cultivares: cultivares representados por pontos mais distantes entre si são mais divergentes que os representados por pontos mais próximos.

Os resultados desta análise permitem fazer inferências sobre a estrutura das populações em estudo, o que constitui informação útil na determinação do número de grupos nas análises de agrupamento.

A dispersão gráfica em relação às duas primeiras variáveis canônicas (Figura 1) permitiu separar, subjetivamente, os 49 cultivares de arroz em sete grupos: dois principais, de maior magnitude; e cinco outros formados por apenas um ou dois cultivares cada um. O grupo I foi formado por cultivares do tipo moderno (classificados inicialmente como pertencentes aos grupos II e III) e dois do tipo tradicional (cultivares T-9 e T-34), os demais grupos foram formados exclusivamente por cultivares do tipo tradicional. Os grupos I e II estão dispostos de maneira bastante próxima, e o grupo II difere do grupo I por apresentar maior média em relação à primeira variável canônica (VC1). Os cultivares dos grupos IV, V, VI e VII são distintamente isolados do resto do material, indicando considerável isolamento e divergência entre esses quatro grupos e o resto do material.

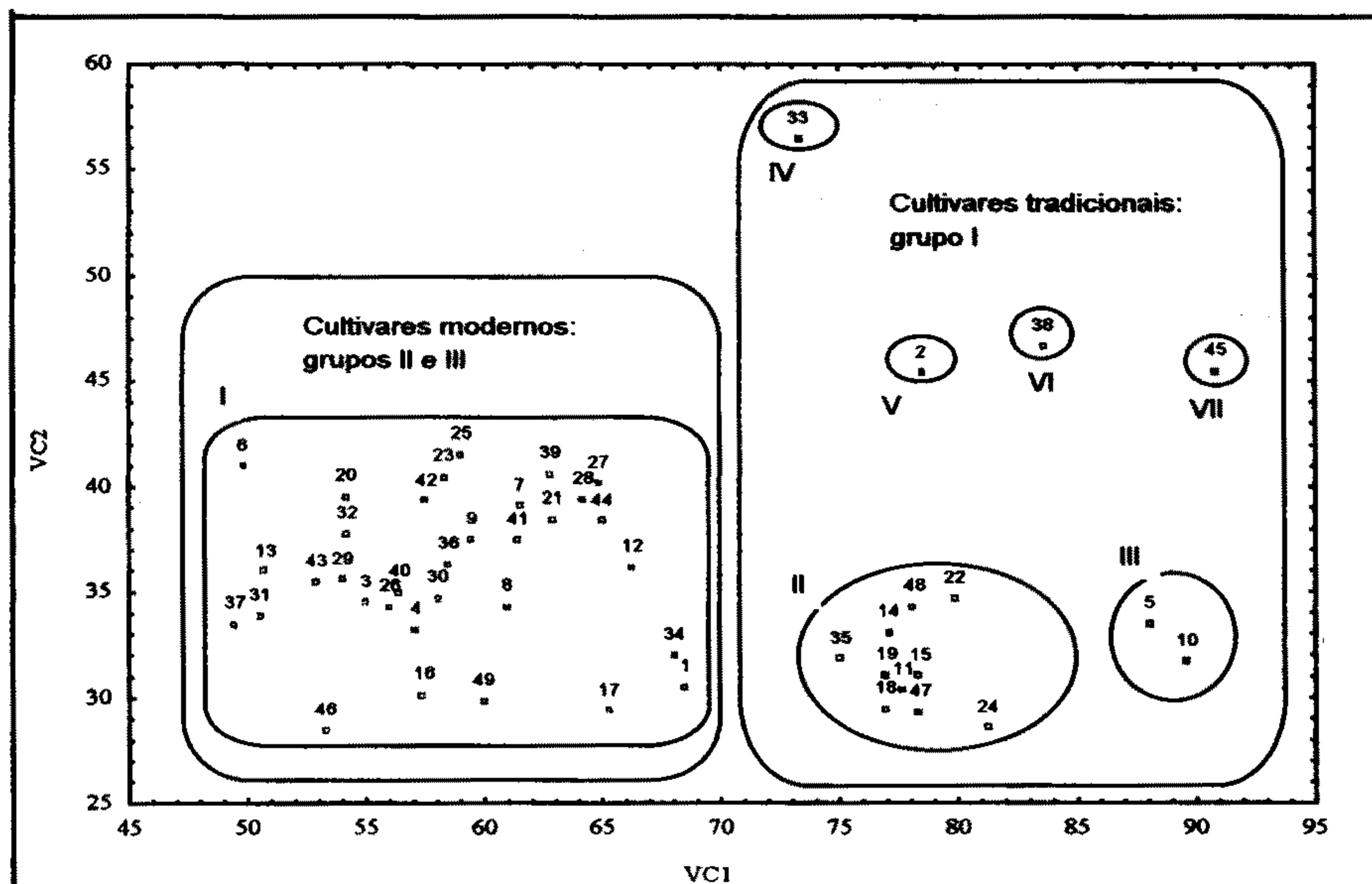
Pela dispersão gráfica, verifica-se que há boa concordância entre os grupos estabelecidos pelos melhoristas, em relação aos padrões moderno e tradicional, e o multivariado, considerando 12 características agronômicas. Entretanto, parece não ser fácil a distinção multivariada dentre os grupos modernos. Outro ponto de destaque diz respeito aos cultivares T-9 e T-34, que, apesar de serem classificados como tradicional, apresentam sua posição relativa próxima dos modernos. Uma avaliação mais criteriosa deve ser considerada em relação a estes cultivares.

Os cultivares tradicionais mais produtivos foram T-18, T-47, T-2 e T-38. Verifica-se que se trata de bons progenitores e que eles ainda exibem certa diversidade genética. Programas de hibridação entre eles podem ser uma alternativa viável para o melhoramento de plantas. De maneira geral, constata-se grande variabilidade no grupo tradicional, com os cultivares mais dispersos.

Constata-se, ainda, que os cultivares modernos do grupo III mais produtivos foram N-25, N-21, N-41 e N-27. Dentre estes, N-21, N-41 e N-27 apresentam certa similaridade, não sendo, portanto, recomendável cruzamentos entre eles. O cultivar N-25 (Cica 8), o mais produtivo, distancia-se destes três outros. O aproveitamento do Cica 8 em programas de hibridação e seu cruzamento com um destes três poderia conduzir a uma população genética superior e com variabilidade a ser explorada em um programa de seleção. Considerando ainda o Cica 8, verifica-se que ele apresenta grande diversidade em relação aos cultivares do grupo II: M-3,

M-4, M-20, M-36, M-1, M-16 e M-6. Dentre estes, merecem destaque M-4 (BR 51-282-8) e M-6 (MR 15), que são também altamente produtivos.

Além de identificar cultivares dissimilares, a análise por variáveis canônicas permite reconhecer aqueles altamente similares, que muitas vezes são réplicas dentro do banco de germoplasma



**FIGURA 1** - Dispersão gráfica em relação às duas primeiras variáveis canônicas, mostrando o relacionamento entre 49 cultivares de arroz em várzea úmida. A identificação dos cultivares está apresentada no Quadro 1.

### *Diversidade genética baseada em análise de agrupamento*

Se o objetivo do programa é aumentar a produtividade, devem-se escolher para cruzamento cultivares de boa performance *per se* e que apresentem maior distância genética, ou que complementem alguma característica de um dos progenitores, pois contribuiriam, em razão de seu não-relacionamento, com o melhor conjunto de genes, os quais, após a recombinação e interação, segregariam, possibilitando assim a obtenção de indivíduos superiores aos pais. Assim, interessa estudar a dissimilaridade entre os cultivares (especialmente entre os de melhor performance), com vistas à seleção de progenitores adequados a um programa de hibridação. Como exemplo, o cultivar mais distante geneticamente em relação ao M-1 (Hn 351-1-8-6), dentre os de melhor performance, com base na distância de Mahalanobis, foi o Desconhecido - T-38. Espera-se que o cruzamento

entre estes cultivares divergentes possibilitem a obtenção de híbridos heteróticos.

A distância generalizada de Mahalanobis é uma medida de distância mais precisa que a euclidiana, pois considera a matriz de variâncias e covariâncias residuais. Assim, os cálculos que consideram esta medida de distância devem ser preferidos.

A análise de agrupamento é uma técnica multivariada essencialmente descritiva e requer cuidados em sua aplicação, devido aos muitos problemas a ela associados (7). O pesquisador precisa decidir quanto à medida de distância ou ao coeficiente de similaridade e à técnica de agrupamento a serem utilizados. Existem diferentes métodos de agrupamento, tanto hierárquicos quanto de otimização, cada um com suas limitações e pressuposições.

Everitt (7) apresentou uma análise comparativa de algumas das técnicas de agrupamento. Entre outras, foram avaliadas as técnicas hierárquicas do vizinho mais próximo, centroid e Ward. O autor relata que, para os dados da pesquisa, das três, a do vizinho mais próximo foi a que apresentou melhor desempenho. Entretanto, afirma o autor, não existe um método que possa ser considerado melhor ou pior em todas as situações; tudo irá depender do tipo de dado com o qual se está trabalhando e da estrutura contida nos dados. Quando estes apresentam grupos esféricos bem separados, o método do vizinho mais próximo caracteriza bem essa estrutura. Contudo, se existirem pontos intermediários, este método deve falhar.

Outros autores defendem ou criticam o método. Lance e Williams (11), citados por Everitt (7), por exemplo, classificam o método do vizinho mais próximo como obsoleto.

Resultados de estudos (1, 2, 3, 4, 5, 8, 9, 12, 15, 16, 17, 18, 20) demonstram a consistência das diferentes técnicas multivariadas e validam a sua aplicação em estudos de divergência genética.

Os resultados da análise hierárquica de agrupamento e de otimização são apresentados nos Quadros 4 e 5. Verificou-se que, utilizando a análise de agrupamento, os cultivares foram reunidos formando grupos mais ou menos homogêneos. Nos métodos hierárquicos, a delimitação de grupos é feita de maneira subjetiva, observando-se os pontos de alta mudança de nível no diagrama de árvore (dendrograma). Por esse motivo, podem-se esperar diferentes padrões de agrupamento, os quais dependem ainda do método de agrupamento e da medida de distância ou de similaridade adotados.

Com relação ao método de Ward, o agrupamento com base na distância de Mahalanobis<sup>1</sup> formou quatro grupos: IWm, IIWm, IIIWm e

---

<sup>1</sup> O método de Ward é baseado na distância euclidiana; porém, neste trabalho, aplicou-se o método também com base na distância de Mahalanobis.

IVWm, e, com base na distância euclidiana, seis: IWe, IIWe, IIIWe, IVWe, VWe e VIWe (Quadros 4 e 5).

Pelo método do vizinho mais próximo (SL), foram formados quatro grupos com base na distância de Mahalanobis (ISLm, IISLm, IIISLm e IVSLm) e sete pela distância euclidiana (ISLe, IISLe, IIISLe, IVSLe, VSLe, VISLe e VIISLe) (Quadros 4 e 5).

Pelo método da ligação média (AL), foram formados quatro grupos pela distância de Mahalanobis (IALm, IIALm, IIIALm e IVALm) e oito pela distância euclidiana (IALe, IIALe, IIIALe, IVALe, VALe, VIALe, VIIALe e VIIIALe) (Quadros 4 e 5).

Pelo método de Tocher (Quadros 4 e 5) formaram-se sete grupos pela distância de Mahalanobis (ITm, IITm, IIITm, IVTm, VTm, VITm e VIITm) e 10 pela distância euclidiana (ITe, IITe, IIITe, IVTe, VTe, VITe, VIITe, VIIITe, IXTe e XTe).

Com base nos resultados da análise de agrupamento, pode-se observar boa concordância entre os padrões de agrupamento dos quatro métodos, tanto em relação à distância de Mahalanobis quanto em relação à distância euclidiana. Em alguns casos, um grupo formado por um dos métodos representa a fusão de dois ou mais grupos formados por outro método de agrupamento.

Comparando os agrupamentos formados utilizando-se a distância de Mahalanobis, observa-se que tanto o método de Ward quanto os do vizinho mais próximo (SL) e da ligação média (AL) separaram os 49 cultivares de arroz em quatro grupos (Quadro 4); o método de Tocher, entretanto, formou sete grupos.

Nos métodos SL e AL, os grupos ISLm e IALm são formados pelos grupos IWm e IIWm menos o indivíduo M-6 (IIWm). O grupo IISLm foi formado pelo grupo IIIWm mais os indivíduos T-10 (IVWm) e T-38 (IVWm); o grupo IIIALm, pelo grupo IIIWm menos o indivíduo T-5 (IIIWm); o grupo IIISLm, pelo indivíduo T-45 (IVWm); o grupo IVSLm, pelo indivíduo M-6 (IIWm); o grupo IIALm, pelo indivíduo M-6 (IIWm); e o grupo IVALm, pelos indivíduos do grupo IVWm mais o indivíduo T-5 (IIIWm).

O grupo ITm foi formado pelos mesmos indivíduos do grupo IWm menos o indivíduo T-34; o grupo IITm, pelos indivíduos do grupo IIIWm menos o indivíduo T-33; o grupo IIITm, pelos indivíduos do grupo IIWm menos o indivíduo M-6; o grupo IVTm, pelos indivíduos do grupo IVWm menos o indivíduo T-45; o grupo VTm, pelos indivíduos T-33 e T-34, dos grupos IIIWm e IWm, respectivamente, os quais estão muito distantes pelo agrupamento de Ward; o grupo VITm, pelo indivíduo M-6 (grupo IIWm); e o grupo VIITm, pelo indivíduo T-45 (grupo IVWm).

Quando o agrupamento foi feito considerando a distância euclidiana, o método de Ward separou os 49 cultivares em seis grupos, o do vizinho

mais próximo (SL) em sete, o da ligação média (AL) em oito e o de Tocher em 10 (Quadro 5). Portanto, não houve concordância quanto ao número de grupos quando se considerou a distância euclidiana.

O grupo ISLe foi igual ao IWe, e o IALe, igual ao IWe menos o indivíduo M-6. O grupo IISLe foi formado pelos grupos IIIWe e VWe; o grupo IIALe, pelo grupo IIWe mais o indivíduo M-6, menos o indivíduo N-25; o grupo IIISLe, pelo grupo IIWe menos o indivíduo N-25; o grupo IIIALe, pelo grupo IIIWe menos o indivíduo M-36; o grupo IVSLe, pelo grupo VIWe mais o grupo IVWe menos os indivíduos T-48 e N-49; e o grupo IVALe, pelo grupo VWe mais o indivíduo M-36 (IIIWe). Os grupos VSLe, VISLe e VIISLe, de apenas um indivíduo, foram formados, respectivamente, pelos indivíduos N-49 (VIWe), T-48 (VIWe) e N-25 (IIWe).

O grupo VALe foi formado pelos indivíduos do grupo IVWe; o grupo VIALe, pelos indivíduos do grupo VIWe menos o indivíduo T-48; e os grupos VIIALe e VIIIALe foram formados, respectivamente, pelos indivíduos T-48 (VIWe) e N-25 (IIWe).

**QUADRO 4 - Padrões de agrupamento de 49 cultivares de arroz em várzea úmida, com base nos métodos de Ward, vizinho mais próximo (SL), ligação média (AL) e Tocher e na distância de Mahalanobis**

Ward	Número de cultivares	Grupos	Cultivares*
	24	I	M-1; N-17; M-16; T-34; M-8; N-27; M-28; N-12; M-44; N-49; M-36; M-3; M-4; N-46; M-23; M-26; N-42; M-20; N-7; N-21; N-41; T-9; N-30; N-40
	9	II	M-6; N-13; N-29; M-37; M-39; N-25; N-31; N-32; M-43
	13	III	T-2; T-18; T-22; T-11; T-24; T-19; T-5; T-14; T-35; T-15; T-47; T-33; T-48
	3	IV	T-10; T-38; T-45
SL	Número de cultivares	Grupos	Cultivares*
	32	I	M-1; N-17; M-16; M-8; N-27; M-28; N-12; M-44; N-49; M-36; M-3; M-4; M-26; N-42; M-23; N-46; M-20; N-7; N-21; N-41; N-30; N-40; T-9; N-13; N-29; M-37; N-25; N-31; N-32; M-43; M-39; T-34
	15	II	T-2; T-11; T-24; T-19; T-22; T-18; T-14; T-35; T-15; T-47; T-48; T-5; T-33; T-10; T-38
	1	III	T-45
	1	IV	M-6

Continua...

## QUADRO 4 - Continuação

AL	Número de cultivares	Grupos	Cultivares*
	32	I	M-1; N-17; M-16; T-34; M-8; N-27; M-28; N-12; M-44; N-49; M-36; M-3; M-4; M-23; M-26; N-42; N-46; M-20; N-7; N-21; N-41; T-9; N-30; N-40; N-13; N-29; M-37; M-39; N-25; N-31; N-32; M-43
	1	II	M-6
	12	III	T-2; T-11; T-24; T-19; T-18; T-22; T-14; T-35; T-15; T-47; T-33; T-48
	4	IV	T-5; T-10; T-38; T-45
Tocher	Número de cultivares	Grupos	Cultivares*
	23	I	N-7; N-21; N-41; N-40; T-9; N-30; M-23; M-3; M-4; M-26; N-46; N-42; M-20; M-28; M-8; N-27; M-36; M-44; N-49; N-12; M-16; N-17; M-1
	12	II	T-14; T-35; T-15; T-22; T-18; T-19; T-24; T-11; T-2; T-47; T-5; T-48
	8	III	N-31; N-32; M-43; N-25; N-13; M-39; M-37; N-29
	2	IV	T-10; T-38
	2	V	T-33; T-34
	1	VI	M-6
	1	VII	T-45
Cultivares marcados com a letra T representam tipos considerados tradicionais, do grupo I; os marcados com a letra M, tipos modernos do grupo II; e os marcados com a letra N, tipos modernos do grupo III.			
* A identificação dos cultivares pode ser feita no Quadro 1.			

**QUADRO 5 - Padrões de agrupamento de 49 cultivares de arroz em várzea úmida, com base nos métodos de Ward, vizinho mais próximo (SL), ligação média (AL) e Tocher e na distância euclidiana média**

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

Continua...

## QUADRO 5 - Continuação

	17	II	T-2; M-20; N-7; N-29; M-3; M-26; N-42; T-9; N-32; N-30; N-13; M-37; N-31; M-23; N-40; M-36; T-38
	3	III	M-4; N-21; M-39
	15	IV	T-5; T-10; T-14; T-45; T-22; T-33; N-46; T-11; T-24; T-15; T-35; T-34; N-17; M-44; T-19
	1	V	N-49
	1	VI	T-48
	1	VII	N-25
AL	Número cultivares	Grupos	Cultivares*
	10	I	M-1; N-12; M-8; M-43; T-47; T-18; N-27; M-28; N-41; M-16
	4	II	M-4; N-21; M-39; M-6
	11	III	T-2; M-20; N-7; N-29; T-38; M-3; M-26; N-42; N-32; T-9; N-30
	6	IV	N-13; M-37; N-31; M-23; N-40; M-36
	6	V	T-5; T-10; T-33; T-14; T-45; T-22
	10	VI	T-11; T-24; T-15; T-35; T-34; N-17; M-44; T-19; N-46; N-49
	1	VII	T-48
	1	VIII	N-25
Tocher	Número de cultivares	Grupos	Cultivares*
	12	I	T-15; T-24; T-11; T-47; T-18; T-10; T-5; T-19; T-35; T-22; M-1; T-14
	22	II	N-7; N-29; M-26; N-32; M-23; M-37; N-40; N-42; N-21; N-31; N-13; M-8; N-12; M-28; N-27; M-3; M-43; M-39; N-41; M-20; N-46; N-17
	2	III	T-38; T-45
	3	IV	T-9; M-36; M-4
	2	V	N-30; N-49
	2	VI	T-2; T-33
	2	VII	M-6; N-25
	2	VIII	T-34; T-48
	1	IX	M-44
	1	X	M-16
Cultivares marcados com a letra T representam tipos considerados tradicionais do grupo I; os marcados com a letra M, tipos modernos do grupo II; e os marcados com a letra N, tipos modernos do grupo III.			
* A identificação dos cultivares pode ser feita no Quadro 1.			



O método de Tocher, neste caso, diverge bastante do método de Ward. O grupo ITe foi formado por indivíduos dos grupos IWe, IVWe e VIWe; o grupo IIe, por indivíduos dos grupos IWe, IIWe, IIIWe, VWe e VIWe; o grupo IIIe, por indivíduos dos grupos IIIWe e IVWe; o grupo IVe, por indivíduos dos grupos IIWe e IIIWe; o grupo Ve, por indivíduos dos grupos IIIWe e VIWe; o grupo VIe, por indivíduos dos grupos IIIWe e IVWe; o grupo VIIe, por indivíduos dos grupos IWe e IIWe; os grupos VIIIe e IXe foram formados por indivíduos do grupo VIWe; e o grupo Xe foi formado por indivíduos do grupo IWe.

Com relação à separação dos indivíduos sabidamente pertencentes aos grupos tradicionais e modernos, todos os quatro métodos aplicados com base na distância de Mahalanobis distinguiram perfeitamente bem estes tipos de cultivares.

Nos agrupamentos feitos com base na distância de Mahalanobis, os grupos IWm e IIWm foram formados por cultivares modernos e os grupos IIIWm e IVWm, por cultivares tradicionais. O grupo IWm agrupou, entretanto, os indivíduos T-34 e T-9, que são tidos como tradicionais. Os grupos ISLm e IVSLm também são formados por cultivares modernos. O grupo ISLm apresentou os indivíduos T-9 e T-34, que são considerados tradicionais. Os grupos IISLm e IIISLm foram formados por cultivares tradicionais.

Os grupos IALm e IIALm foram formados por cultivares modernos, com exceção dos indivíduos T-9 e T-34, no grupo IALm, que são cultivares tradicionais. Os grupos IIIALm e IVALm são cultivares tradicionais.

Os grupos ITm, IIITm e VITm foram formados por cultivares modernos. ITm agrupou o indivíduo T-9, que pertence aos cultivares tradicionais. Os grupos IIITm, IVITm, VITm e VIITm foram formados por cultivares tradicionais.

Com relação à matriz de distância euclidiana, a separação entre cultivares do tipo tradicional e os do tipo moderno não foi tão satisfatória quanto o foi quando a análise de agrupamento foi aplicada sobre a matriz de distância de Mahalanobis. Nos agrupamentos com base na distância euclidiana, os grupos IWe, IIWe, IIIWe e VWe foram formados por cultivares modernos, e os grupos IVWe e VIWe, por cultivares tradicionais.

Os grupos ISLe, IISLe, IIISLe, VSLe e VIISLe foram formados por cultivares modernos, e os grupos IVSLe e VISLe, por cultivares tradicionais. Os grupos IALe, IIALe, IIIALe, IVALe e VIILe foram formados por cultivares modernos, e os grupos VALe, VIILe e VIIILe, por cultivares tradicionais. Os grupos IWe, ISLe e IALe agruparam os indivíduos T-18 e T-47, que são cultivares tradicionais. Os grupos IIIWe, IISLe e IIIALe agruparam os indivíduos T-2, T-9 e T-38, que são

cultivares tradicionais; os grupos VIWe e VIALe agruparam os indivíduos N-17, M-44, N-46 e N-49, que são cultivares modernos; e o grupo IVSLe agrupou os indivíduos N-17, M-44 e N-46, que são cultivares modernos. Os grupos II<sup>Te</sup>, IV<sup>Te</sup> (com exceção do indivíduo T-9), V<sup>Te</sup>, VII<sup>Te</sup>, XI<sup>Te</sup> e X<sup>Te</sup> foram formados por cultivares modernos, e os grupos I<sup>Te</sup> (com exceção do indivíduo M-1), III<sup>Te</sup>, VI<sup>Te</sup> e VIII<sup>Te</sup>, por cultivares tradicionais.

Considerando a distância de Mahalanobis, dentre os cultivares de melhor performance em termos de produção de grãos, as combinações mais adequadas, indicadas simultaneamente pelos métodos de Ward, da ligação média e de Tocher, foram: T-38 e M-43; T-38 e N-32; T-38 e N-25; T-38 e M-39; e T-38 e N-29.

Os métodos de Tocher e da ligação média indicaram exatamente as mesmas combinações quando foi utilizada a distância de Mahalanobis, e o método do vizinho mais próximo indicou as combinações: M-6 e T-38; M-6 e T-47; M-6 e T-18; e M-6 e T-2.

Considerando-se a distância euclidiana, os resultados não foram muito coerentes em todos os métodos. No método do vizinho mais próximo e no da ligação média foram indicadas as combinações N-25 e M-4; N-25 e N-21; e N-25 e M-39 como as mais divergentes. Entretanto, o método de Ward reuniu os cultivares N-25, M-4, N-21 e M-39 no mesmo grupo, indicando estreito relacionamento genético entre eles. Os métodos da ligação média e do vizinho mais próximo com base na distância de Mahalanobis também agruparam estes cultivares num mesmo grupo.

Os métodos de Tocher e de Ward, com base na distância euclidiana, indicaram as combinações N-25 e T-2 e N-25 e T-38 como as mais divergentes.

Em face do exposto, conclui-se que tanto o padrão de agrupamento quanto as combinações indicadas para cruzamento são consistentes apenas quando os métodos são utilizados com a mesma medida de distância.

Utilizou-se também a análise de agrupamento para identificar os acessos mais similares, com vistas à simplificação de banco de germoplasma. Grupos de cultivares que apresentam as menores distâncias inter e intragrupais são mais estreitamente relacionados. Assim, podem-se selecionar alguns desses cultivares para conservação no banco de germoplasma, descartando-se os demais.

Análises de agrupamento executadas sobre os dados de distância de Mahalanobis (estatística  $D^2$ ) identificaram pelo método de Ward os grupos I e II (ambos compostos de cultivares do tipo moderno) como os mais similares, pois apresentam a menor distância intergrupais. A segunda menor distância intergrupais foi verificada entre os grupos III e IV (formados por cultivares tradicionais).

No método do vizinho mais próximo, os grupos mais similares foram I e IV (cultivares modernos) e I e III (um formado por cultivares modernos e o outro por um cultivar tradicional); no método da ligação média, as menores distâncias intergrupais foram identificadas nos grupos III e IV (cultivares tradicionais) e I e II (cultivares modernos); e o método de Tocher identificou os grupos I e III (cultivares modernos) e I e VI (cultivares modernos) como os mais similares.

Com base na distância euclidiana, as menores distâncias intergrupais foram observadas nos grupos III e V (cultivares modernos) e IV e VI (cultivares tradicionais), obtidos pelo método de Ward; entre os grupos IV e V (um formado por cultivares tradicionais e o outro por um cultivar moderno) e I e III (cultivares modernos), pelo método do vizinho mais próximo; entre os grupos III e IV (cultivares modernos) e I e II (cultivares modernos), pelo método da ligação média; e entre os grupos II e IV (cultivares modernos) e V e IX, pelo método de Tocher.

Os acessos consistentemente similares nos quatro métodos obtidos com o uso da distância de Mahalanobis foram: M-1, M-3, M-4, N-7, M-8, T-9, N-12, N-13, M-16, N-17, M-20, N-21, M-23, N-25, M-26, N-27, M-28, N-29, N-30, N-31, N-32, M-36, M-37, M-39, N-40, N-41, N-42, M-43, M-44, N-46 e N-49.

Quando se utilizou distância euclidiana, os acessos mais similares foram: M-3, N-7, M-8, T-9, N-12, N-13, N-17, M-20, N-21, M-23, M-26, N-27, M-28, N-29, N-30, N-31, N-32, M-36, M-37, M-39, N-40, N-41, N-42, M-43, M-44, N-46 e N-49.

## CONCLUSÕES

1) Os cultivares mais promissores em produção de grãos são: Lageado, Come Cru, De Abril, Desconhecido e C 1016-1, dentro do grupo tradicional (grupo I); no grupo II, cultivares modernos, os mais promissores são: CR 1002 (M-39), BR 51-282-8 (M-4), BR 51-46-5 (M-43), MR 15 (M-6) e Hn 351-1-8-6 (M-1); e no grupo III, cultivares modernos, os mais promissores são: Cica 8 (N-25), IAC 899 (N-21), IET 5389 (N-29), Cica 9 (N-32) e BR 1 (N-41).

2) As diferentes medidas de distância (euclidiana e de Mahalanobis) e os métodos de agrupamento (métodos de Ward, do vizinho mais próximo, da ligação média e de Tocher) conduzem a diferentes padrões de agrupamento.

3) Apesar da diferença do número de grupos e do padrão de agrupamento, de maneira geral os resultados são concordantes.

4) Tanto o padrão de agrupamento quanto as combinações indicadas para cruzamento são mais consistentes quando os métodos são utilizados com a mesma medida de distância.

## REFERÊNCIAS

1. ADAMS, M.W. An estimation of homogeneity in crop plants, with special reference to genetic vulnerability in the dry bean, *Phaseolus vulgaris*. L. *Euphytica*, 26:665-79, 1977.
2. ALLISON, D.B. & HESHKA, S. Toward an empirically derived typology of obese persons: derivation in a nonclinical sample. *International Journal of Eating Disorders*, 13:93-108, 1993.
3. BHATT, G.M. An application of multivariate analysis to selection for quality characters in wheat. *Australian Journal of Agricultural Research*, 27:11-8, 1976.
4. BRITZ, T.J. & RIEDEL, K.H.J. Propionibacterium species diversity in Leerdammer cheese. *International Journal of Food Microbiology*, 22:257-67, 1994.
5. BROWN, J.S. Principal component and cluster analyses of cotton cultivar variability across the U.S. Cotton Belt. *Crop Science*, 31:915-22, 1991.
6. CRUZ, C.D. Programa GENES – Versão Windows: aplicativo computacional em genética e estatística. Viçosa, Editora UFV, 2001. 648p.
7. EVERITT, B. Cluster analysis. London, Heinemann Educational Books, 1977. 122 p.
8. GHADERI, A.; ADAMS, M.W. & NASSIB, A.M. Relationship between genetic distance and heterosis for yield and morphological traits in dry edible bean and faba bean. *Crop Science*, 24:37-42, 1984.
9. GOWER, J.C. & ROSS, G.J.S. Minimum spanning trees and single linkage cluster analysis. *Applied Statistics*, 18:54-64, 1969.
10. IEZZONI, A.F. & PRITTS, M.P. Applications of principal component analysis to horticultural research. *HortScience*, 26:334-8, 1991.
11. LANCE, G.N. & WILLIAMS, W.T. A general theory of classificatory sorting strategies. *Computer Journal*, 1:373-80, 1967.
12. MALUF, W.R. & FERREIRA, P.E. Análise multivariada da divergência genética em feijão-vagem (*Phaseolus vulgaris* L.). *Horticultura Brasileira*, 1:31-4, 1983.
13. MARDIA, K.V.; KENT, J.T. & BIBBY, J.M. Multivariate analysis. London, Academic, 1979. 521 p.
14. RANGEL, P.H.N. & NEVES, P.C.F. Seleção recorrente em arroz irrigado no Brasil. In: Taller Internacional sobre Seleção Recorrente em Arroz, 1, 1995, Goiânia. Anais, Goiânia, CNPAF/EMBRAPA, 1995. p. 114-28.
15. ROSS, G.J.S. Single linkage cluster analysis. *Applied Statistics*, 18:106-10, 1969.
16. SINGH, R.S. Genetic divergence in indigenous varieties of rice grown in Mirzapur district, U.P. *International Rice Research Notes*, 6(2):3-4, 1981.
17. SINGH, Y.P.; KUMAR, A. & CHAUHAN, B.P.S. Genetic divergence in pearl millet. *Indian Journal of Genetics & Plant Breeding*, 41:186-90, 1981.
18. SMITH, S.E.; AL-DOSS, A. & WARBURTON, M. Morphological and agronomic variation in North African and Arabian alfalfas. *Crop Science*, 31:1159-63, 1991.
19. SNEATH, P.H.A. & SOKAL, R.R. Numerical taxonomy. San Francisco, W.H. Freeman, 1973. 573 p.
20. VEN, W.T.G. VAN-DE.; DUNCAN, N.; RAMSAY, G.; PHILLIPS, M.; POWELL, W. & WAUGH, R. Taxonomic relationships between *V. faba* and its relatives based on nuclear and mitochondrial RFLPs and PCR analysis. *Theoretical and Applied Genetics*, 86:71-80, 1993.