DETERMINAÇÃO DE DIVERGÊNCIA GENÉTICA EM GERMOPLASMAS DE ARROZ
(*Oryza sativa* L.) ATRAVÉS DE ANÁLISES ELETROFORÉTICAS
DE PROTEÍNAS DE GRÃO

RICARDO MONTALVAN DEL AGUILA

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, para obtenção do título de Mestre em Agronomia. Área de concentração: Genética e Melhoramento de Plantas.

P I R A C I C A B A
Estado de São Paulo - Brasil
Novembro - 1990

Ficha catalográfica preparada pela Seção de Livros da
Divisão de Biblioteca e Documentação - PCAP/USP

Montalvan del Aguila, Ricardo

M763d Determinação de divergência genética em germoplasmas de arroz (Oryza sativa L.) através de análises eletroforéticas de proteínas de grão. Piracicaba, 1990.

80p. ilus.

Diss. (Mestre) -- ESALQ
Bibliografia.

1. Análise multivariada 2. Arroz - Genética 3. Arroz - Germoplasma - Divergência genética 4. Eletroforese 5. Proteína do grão de arroz I. Escola Superior de Agricultura Luiz de Queiroz, Piracicaba

CDD 633.18

DETERMINAÇÃO DE DIVERGÊNCIA GENÉTICA EM GERMOPLASMAS DE ARROZ
(*Oryza sativa* L.) ATRAVÉS DE ANÁLISES ELETROFORÉTICAS
DE PROTEÍNAS DE GRÃO

RICARDO MONTALVAN DEL AGUILA

Engenheiro-Agrônomo

Orientador: PROF. DR. AKIHIKO ANDO

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, para obtenção do título de Mestre em Agronomia. Área de concentração: Genética e Melhoramento de Plantas.

P I R A C I C A B A
Estado de São Paulo - Brasil
Novembro - 1990

DETERMINAÇÃO DE DIVERGÊNCIA GENÉTICA EM GERMOPLASMAS DE ARROZ
(*Oryza sativa* L.) ATRAVÉS DE ANÁLISES ELETROFORÉTICAS
DE PROTEÍNAS DE GRÃO

RICARDO MONTALVAN DEL AGUILA

Aprovada em: 19.12.90

Comissão julgadora:

Prof. Dr. Akihiko Ando	ESALQ/USP
Dr. Eric Derbyshire	CENA/USP
Dr. Sergio Echeverrigaray Laguna	ESALQ/USP

PROF. DR. AKIHIKO ANDO

Orientador

A meus pais, Hortencia e José Enrique (in memoriam)

A minha esposa Maria Alicia

A meus filhos e irmãos

DEDICO

AGRADECIMENTOS

Este trabalho é produto do esforço, dedicação e colaboração de muitas pessoas e instituições. A todos eles meu agradecimento sincero. Por razões de espaço serão mencionados alguns deles, com o perdão para os omitidos;

- Ao Dr. Akihiko Ando, pela sua excelente direção e constante preocupação na execução do trabalho e sobretudo pela atitude sempre amigável e compreensiva;
- Ao Dr. Sergio Echeverrigaray Laguna, pela sua contribuição decisiva na execução prática neste trabalho, e pela permanente disposição de ajuda e amizade;
- Ao Dr. Joaquim Aparecido Machado e a "Sementes Agroceres" pela facilidades oferecidas no seu laboratório de Santa Cruz das Palmeiras;
- Ao Dr. Cassio Godoy, pela assessoria e valiosa ajuda nos cálculos estatísticos;
- Ao Dr. Luis A. Lopes, pela amizade sincera e seus ensinamentos em Estatística Multivariada;
- Ao Instituto de Pesquisa da Amazonia Peruana (IIAP), nas pessoas dos engenheiros Roger Beauzeville Zumaeta e Wilfredo Meza Aguilar que possibilitaram a minha participação no curso de pós-graduação;

- A CAPES, pela bolsa concedida;
- Aos docentes do Departamento de Genética pelos ensinamentos, e aos colegas, especialmente a Hermes Aramendis, Roberto Carbonera, Braz Pacova e Dirce Carraro pela sua grande amizade.
- Aos funcionários do Departamento de Genética, com ênfase em especial para as Sras. Nilce Puga e Terezinha Barrete da biblioteca, a Sra. Carmen F. Piloto e Humberto Gomez por inúmeros favores.

ÍNDICE

	Página
LISTA DE FIGURAS	vii.
LISTA DE TABELAS	viii.
RESUMO	ix.
SUMMARY	xi.
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1. Eletroforese de Proteínas Desnaturadas em Gel de Poliacrilamida (SDS-PAGE)	3
2.2. Classificação e Descriptores Bioquímicos	5
2.3. Divergência Genética	6
2.3.1. Medição da distância genética	7
2.3.2. Idéia de distância estatística	8
2.3.3. Distância Euclidiana e de Mahalanobis ..	9
2.4. Distâncias entre Populações Caracterizadas por Atributos Qualitativos	11
2.5. Relações Genéticas a Partir da Avaliação Quan- titativa de Bandas Eletroforéticas de Proteí- nas de Semente	12
2.6. Análise de Conglomeração	16
2.7. Análise de Função Discriminante	18
3. MATERIAL E MÉTODOS	21
3.1. Materiais	21
3.2. Métodos	24
3.2.1. Trabalhos no campo	24
3.2.2. Trabalhos no laboratório	24

	Página
3.2.2.1. Extração da proteína solúvel...	24
3.2.2.2. Quantificação de proteína	25
3.2.2.3. Preparação da amostra desnatura- rada	26
3.2.2.4. Eletroforese SDS-PAGE	27
3.2.2.5. Densitometria	29
3.3. Metodologias Estatísticas	30
3.3.1. Análise univariado	30
3.3.2. Análise multivariada	31
3.3.2.1. Agrupamento das variedades ...	32
3.3.2.2. Análise de função discriminan- te (AFD)	34
4. RESULTADOS E DISCUSSÃO	39
4.1. Análise Eletroforética e Univariada	39
4.2. Distância Entre Variedades e Agrupamento	47
4.3. Diferenciação com Análise de Função Discrimi- nante	61
4.3.1. Discriminação Entre Variedades Melhora- das e Não Melhoradas	65
4.3.2. Discriminação Entre Variedades Brasi- leiras e Japonesas	70
5. CONCLUSÕES	74
REFERENCIAS BIBLIOGRAFICAS	76

LISTA DE FIGURAS

Figura	Página
1 Um típico gel SDS-PAGE e traços densitométricos de proteínas de cariopse de arroz (variedade IAC-5544).	40
2 Gel característico com os padrões eletroforéticos de 12 variedades.	41
3 Dendograma representativo entre 58 variedades brasileiras de arroz, obtido pelo método do Vizinho mais Próximo.	58.
4 Dispersão de escores discriminantes de 58 variedades brasileiras de arroz: a) escores da 1 componente; b) escores da 2 componente; c) escores da 1 e 2 componentes.	62.
5 Dispersão de escores discriminantes de 58 variedades brasileiras de arroz para o contraste de variedades melhoradas e não melhoradas.	68.
6 Dispersão de escores discriminantes de 67 variedades de arroz para o contraste de variedades brasileiras versus japonesas.	71.

LISTA DE TABELAS

Tabela	Página
1 Algumas características das variedades brasileiras e japonesas (J).	21
2 Composição de soluções para quantificação de proteínas pelo método de Bradford (1976)	25
3 Composição de soluções estoques do procedimento eletroforético SDS-PAGE	26
4 Composição dos géis (SDS-PAGE) (roteiro de preparação)	28
5 Padrões eletroforéticos (SDS-PAGE) de 67 variedades de arroz mostrando 16 frações proteicas (variáveis) com os seus respectivos pesos moleculares (*) (em kd) e valores corrigidos de concentração (em mm) medidas no densitômetro.	42
6 Análise de variância de 16 caracteres proteicos em 58 variedades brasileiras de arroz, no delineamento de blocos ao acaso. Piracicaba, 1990.	45
7 Distância Euclidiana média entre 58 germoplasmas brasileiras de arroz (<i>Oryza sativa</i> L.).	48
8 Composição de grupos de 58 variedades brasileiras de arroz, pelo método de Tocher. Piracicaba, 1990.	54.
9 As primeiras 4 componentes da análise discriminante para a classificação de 58 variedades brasileiras de arroz. Piracicaba, 1990.	64.

DETERMINAÇÃO DE DIVERGÊNCIA GENÉTICA EM GERMOPLASMAS DE ARROZ
(*Oryza sativa* L.) ATRAVÉS DE ANÁLISES ELETROFORÉTICAS
(SDS-PAGE) DE PROTEÍNAS DE GRÃO

Autor: RICARDO MONTALVAN DEL AGUILA

Orientador: PROF. DR. AKIHIKO ANDO

RESUMO

Visando caracterizar e aglomerar variedades em grupos de diversidade, foram avaliadas as variações quantitativas nas proteínas de grão, separadas por SDS-PAGE, entre 67 germoplasmas de arroz (*Oryza sativa* L.) (59 brasileiros e 8 japoneses). A avaliação quantitativa foi desenvolvida através de análise densitométrica dos padrões eletroforéticos. Os dados obtidos permitiram a determinação da distância genética entre pares de variedades, utilizando-se a distância euclidiana média. A análise de conglomeração pelo método de Tocher possibilitou a formação de 10 grupos de diversidade, enquanto que pelo método do vizinho mais próximo foi obtido um dendograma das relações de afinidade entre as variedades. Foi observado, em ambos os métodos de agrupamento, que em termos gerais, o gradiente de diferenciação está relacionado com a procedência dos materiais e com o nível de melhoramento das variedades. Foi aplicada a análise

de função discriminante, que em gráficos de dispersão confirmou que é possível uma clara discriminação entre variedades brasileiras e japonesas. Entretanto, a diferenciação entre variedades melhoradas (M) e não melhoradas (O) não foi tão manifesta, embora tenha sido possível estabelecer faixas para M e O e uma faixa de transição entre elas.

As técnicas estatísticas aplicadas aos componentes proteicos mencionados indicam que o conhecimento das proteínas de semente permite a avaliação da afinidade genética entre variedades e facilita a formação de grupos de diversidade e a diferenciação entre variedades de "backgrounds" genéticos similares.

GENETIC DIVERGENCE AND CLASSIFICATION OF GERMPLASMS OF RICE
(*Oryza sativa* L.) THROUGH ELECTROPHORETIC TECHNIQS
(SDS-PAGE) OF SEED'S PROTEINS

Author: RICARDO MONTALVAN DEL AGUILA

Advisir: PROF. DR. AKIHIKO ANDO

SUMMARY

Variation among 67 rice cultivars (*Oryza sativa* L.) was evaluated comparing their soluble grain proteins profiles on SDS-PAGE gels. The protein fractions were quantitatively analysed by densitometric scanning of the gels. Utilization of this methodology allowed an evaluation of genetic distances among pairs of cultivars, by estimation of the mean euclidian distances. Ten groups of diversity were formed by of the Tocher's method of conglomeration analysis, which were confirmed by the dendogram of afinity relations (single linkage method). In general, the results showed that differentiation grades were related to the origin of the materials and their level of improvement. Graphic analysis of discriminant function showed a clear discrimination between brazilian and japanese cultivars. However, differenciacion between improved (MD and

non improved (O) cultivars was not so clear, showing a transition range between them.

Statistical techniques applied to protein components indicate that knowledge of seed proteins allows to evaluate the genetic affinity among varieties and enables to form diversity groups and to discriminate them with similar genetic background.

1. INTRODUÇÃO

O arroz cultivado (*Oryza sativa* L.) foi introduzido no Brasil no século XVIII. Desde essa época, numerosas variedades foram trazidas por imigrantes de vários países do mundo. Em conjunto, são milhares de variedades espalhadas por todo o país, constituindo-se em bancos de germoplasmas brasileiros desta cultura, em várias instituições.

Paradoxalmente, no entanto, a utilização destes germoplasmas, por parte de programas de melhoramento é mínima, até mesmo irrisória. Uma das causas deste uso incipiente, segundo van Sloten citado por VALLS (1988), é a falta de informação sobre os materiais dos bancos, ou mesmo quando disponível, tal informação é inadequada ou insuficiente para o melhorista.

O conhecimento da diversidade genética entre parentais é considerado de suma importância na obtenção de segregantes transgressivos e subsequente desenvolvimento de variedades melhoradas (GADHERY *et alii*, 1984). Por conseguinte, o conhecimento prévio de grupos de divergência genética em uma cultura é muito útil para diminuir o número de entradas nos estágios iniciais ("screening" e cruzamento) no programa de melhoramento.

A caracterização e avaliação de germoplasma - nos moldes tradicionais - é muito onerosa. A demanda de informação dos melhoristas, sobre grupos de divergência, entre outros tópicos, pode esperar um período longo. Por outro lado, a quantidade e níveis de diversidade genética de numerosas culturas têm sido estudados na base de caracteres isoenzimáticos.

A eletroforese em géis de poliacrilamida é um método de fracionamento molecular vantajoso quanto à sensibilidade e alto poder de resolução, combinados à economia e simplicidade (GUIMARAES, 1978). Esta técnica, aplicada a proteínas de semente, também tem sido usada com sucesso, para estabelecer relações taxonômicas em grande número de espécies de plantas (LADIZINSKY & HYMOWITZ, 1979).

O uso de eletroforese em proteínas de semente, que supera em simplicidade ao uso de isoenzimas, é considerado útil para revelar associações genéticas de divergência, que possam ser usadas em programas de melhoramento.

O presente trabalho foi feito para caracterizar e aglomerar variedades de arroz em grupos de diversidade. Usou-se para isso, dados quantitativos de densitogramas de bandas de proteínas solúveis de semente, resultantes da eletroforese em géis de poliacrilamida com o uso do dissociante sódio duodecil sulfato (SDS-PAGE). Tais dados foram submetidos a técnicas de análise multivariada de utilidade já bem reconhecida, como são, análise de agrupamento e análise discriminante.

2. REVISÃO DE LITERATURA

2.1. Eletroforese de proteínas desnaturadas em gel de poliacrilamida (SDS-PAGE)

Eletroforese é uma técnica que permite a separação de macromoléculas eletricamente carregadas, numa solução quando a esta é aplicada uma corrente elétrica. O método está baseado no fato de que moléculas tais como DNA, RNA e proteínas possuem uma carga e são capazes de se movimentar quando colocadas em um campo elétrico. As primeiras aplicações de métodos eletroforéticos foram realizadas em soluções de sacarose (COOPER, 1977), embora a inconveniência desses métodos e o equipamento caro necessário tenham limitado severamente sua aceitação. O uso de géis de amido em lugar de sacarose, assim como suportes e agentes de anticonvecção, aumentaram a aplicação de eletroforese a problemas biológicos; mas foi o advento dos géis de acrilamida que impulsionou este método à sua atual popularidade.

Diversos sistemas de eletroforese foram desenvolvidos e têm contribuído enormemente para estudos de caracterização de espécies ou variedades (FORDE *et alii*, 1980; WEN & LUTHE, 1985, etc.). Dentro dos sistemas eletro-

foréticos mais utilizados, a eletroforese de proteínas com o uso do dissociante SDS em gel de poliacrilamida (SDS-PAGE) ocupa um lugar relevante nos trabalhos de caracterização.

Foi mostrado que o sódio duodecil sulfato (SDS), um detergente aniônico, liga as regiões hidrofóbicas de proteínas para separar a maioria em suas unidades componentes (polipeptídeos). Assim, as proteínas perdem o efeito de carga específico, separando-se devido a suas diferenças de pesos moleculares. Esta separação permite uma melhor resolução das bandas, tornando possível a avaliação, a nível qualitativo (ausência/presença) e quantitativo (diferenças de concentração) dos polipeptídeos existentes numa amostra (CARRARO, 1990). PANDA *et alii* (1986) também salientam o PAGE como uma poderosa técnica para avaliar homologia gênica ao nível molecular devido a sua superior capacidade para resolução de componentes e acrescentam uma outra característica: o PAGE fornece uma ferramenta adicional para identificação e delimitação de espécies e tem sido particularmente útil na dedução de relações sistemáticas entre grupos onde dados citológicos ou morfológicos não dão resultado.

GUIMARAES (1978) mostra algumas características gerais dos géis de poliacrilamida, que são relatadas a seguir: "As moléculas de acrilamida monomérica são postas a polimerizarem-se em mistura com pequena quantidade de bis-acrilamida, sob a ação iniciadora catalítica do íon persulfato, ativado pela amina terciária N.N.N'.N'.-tetrametil etilanolodiamina (TEMED).

O polímero de acrilamida forma longas moléculas lineares, mas a incorporação, em intervalos regulares, da bis-acrilamida, introduz ligações cruzadas entre os polímeros. As distâncias entre as ligações cruzadas são determinadas pela concentração da bis-acrilamida, em relação à quantidade total de monômeros, também chamada de concentração do gel.

2.2. Classificação e descriptors bioquímicos

A classificação, segundo Everitt, citado por CURI (1983), pode ser entendida como um processo para localizar entidades em classes inicialmente indefinidas, de modo que indivíduos da mesma classe sejam similares entre si, definidos pelas variáveis consideradas. Essas classes de indivíduos semelhantes são os agrupamentos.

O termo variáveis, utilizado por Everitt, resulta mais apropriado dentro do contexto do presente trabalho, como "descriptores"¹. Para fins de classificação, na abordagem do melhoramento, é fundamental a escolha de um jogo de descriptores convenientes. Estes descriptores devem possuir suficiente capacidade de discriminação entre as variedades a avaliar; e revelar, por sua vez, as relações genéticas de semelhança entre os mesmos. A descrição exata de variedades de arroz cultivadas, segundo ALIAGA-MOREL

¹ Tradução do inglês "descriptors", utilizado por VALLS (1988).

(1987) é dificultada pela grande similaridade morfológica, por causa da forte relação genética entre elas. Além disto deve-se considerar que o uso de grande número de caracteres morfológicos é, com frequência, insuficiente para distinguir algumas variedades.

Por outro lado o pó de semente é relativamente fácil de se manejar na extração de proteína e, mais ainda, as sementes maduras podem ser consideradas um estágio relativamente estável, em especial quanto a sua composição protéica (CLADIZINSKY & HYMOWITZ, 1979). Mesmo quando condições ambientais durante o desenvolvimento possam levar a alterações na composição das sementes, as amostras obtidas sob essas condições representam uma fonte apropriada a estudos de distanciamento. Entretanto, é bom deixar claro que os caracteres bioquímicos complementam os descritores de natureza morfológica, agronômica, etc., que são de indiscutível valor para o melhorista, assegurando mais confiabilidade e estabilidade. E mais ainda, os caracteres morfológicos e agronômicos não são simultâneos no tempo, e a sua colheita requer muitas vezes um prolongado período de tempo.

2.3. Divergência genética

A divergência genética é nada mais que a expressão da dissimilaridade entre dois indivíduos ou populações. A quantificação dessa dessemelhança é feita, geralmente, através da distância estatística que, quando aplicada à

genética, é chamada distância genética.

A ênfase da divergência está relacionada ao grau de diversidade dos parentais, como um importante fator que possa contribuir para a obtenção de heterose de alto rendimento ou para segregação transgressiva.

2.3.1. Medição da distância genética

A estimação da distância genética entre populações pode ser feita com base:

- a) no estudo de seus ancestrais;
- b) na diversidade de origem (divergência geográfica) e
- c) na sua composição genética (GHADERI *et alii*, 1984).

O fato de que indivíduos aparentados se assemelhem uns aos outros é indicativo de sua relação genética. Segundo LEFORT-BUSON (1985), as distâncias baseadas em coeficientes de identidade ou de parentesco representam uma informação condensada, mais ou menos precisa, de genealogia desses indivíduos, embora estejam na hipótese de ausência de seleção gamética e zigótica, e de seleção natural e artificial. Cabe salientar que a distância baseada em dados genealógicos só é possível entre indivíduos com algum grau de parentesco, e que não pode ser aplicada a grande parte das comparações entre germoplasmas usados em programas de melhoramento. Deve ser também considerada que a dessemelhança genética entre dois indivíduos aparentados não é de 100% como indicam os dados genealógicos.

Na ausência de outras informações, a diversidade de origem dos parentais tem sido usada como uma medida de sua divergência, embora este critério tenha exceções (OLIVEIRA, 1989). Não por acaso, MURTY & ARUNACHALAM (1966) não encontraram paralelismo entre diversidade genética e distribuição geográfica. Segundo estes autores, a deriva genética e a seleção em diferentes ambientes podem causar maior diversidade que a distância geográfica.

A distância genética baseada na composição genética de populações pode ser vista sobre a base de frequência de diferentes genótipos (distância genotípica) ou frequência de vários alelos em um dado loco (distância gênica) (GHADERI *et alii*, 1984).

2.3.2. Idéia de distância estatística

O conceito de distância estatística, na sua abordagem mais simples, pode ser visto em literatura especializada de estatística (MANLY, 1986) de biologia (GOODMAN, 1972), ou até mesmo de genética (JACQUARD, 1974; ARUNACHALAM, 1981; LEFORT-BUSON, 1985). Segundo estes autores, quando duas entidades (objetos, indivíduos, populações, etc.), forem totalmente tipificadas por uma simples característica, então será fácil ver que a diferença entre os valores desse caráter corresponde à melhor medida de sua distância. Considerando-se dois genótipos G_1 e G_2 , os quais podem ser perfeitamente individualizados por duas características

independentes x_1 e x_2 , pode-se adotar a convenção de que x_{ij} denota o valor do genótipo i no caráter j . Neste caso, o genótipo G_1 é caracterizado por (x_{11}, x_{12}) e o genótipo G_2 por (x_{21}, x_{22}) . Nestas condições a distância entre G_1 e G_2 pode ser obtida de $D^2 = (x_{21} - x_{11})^2 + (x_{22} - x_{12})^2$. Se forem consideradas as diferenças nos valores dos caracteres dos genótipos G_1 e G_2 como sendo d_1 e d_2 , ter-se-á:

$$d^2 = (x_{21} - x_{11})^2 + (x_{22} - x_{12})^2$$

$$d_1 = x_{21} - x_{11}$$

$$d_2 = x_{22} - x_{12}$$

tal que,

$$d^2 = d_1^2 + d_2^2$$

para j caracteres ter-se-á: $d^2 = \sum_{j=1}^J d_j^2$.

Têm sido proposta muitas medidas de distância nas décadas passadas, atendendo a vários objetivos (LEFORT-BUSON, 1985 e JACQUARD, 1974). A distância generalizada de Mahalanobis (MAHALANOBIS, 1936) ocupa um lugar único no melhoramento de plantas, de acordo com a expressão de ARUNACHALAM (1981).

2.3.3. Distância Euclidiana e de Mahalanobis

É quase obrigatório, antes de tratar de distância de Mahalanobis, referir-se a distância euclidiana. O conceito de distância euclidiana já foi esboçado no parágrafo

fo onde se dá a idéia de distância. Sua fórmula é:

$$\text{distância euclidiana} = \left[\sum_{j=1}^J (x_{ij} - x_{i'j})^2 \right]^{1/2}$$

onde:

x_{ij} = o valor do caracter j no indivíduo (ou população) i ;

$x_{i'j}$ = o valor do caracter j no indivíduo (ou população) i' .

A distância euclidiana tende a aumentar com o número p de caracteres anotados. Este inconveniente é superado com o uso de uma modificação, a chamada distância euclidiana média:

$$\text{distância euclidiana média} = \frac{1}{\sqrt{p}} (\text{distância euclidiana})^{1/2}$$

As distâncias euclidiana e euclidiana média são de cálculo fácil, mas têm algumas limitações. A principal delas é aquela que assume que os caracteres avaliados são independentes, o que nem sempre é verdade. Por sua vez, a distância generalizada de Mahalanobis não tem essa restrição, pois ela libera os dados de prováveis correlações, no cálculo da matriz comum de covariâncias para os indivíduos ou populações. Fora dessa importante qualidade, a distância euclidiana ou euclidiana média é semelhante à distância de Mahalanobis. Esta última é expressa por:

$$D_{(ij)}^2 = (x_i - x_j)' S^{-1} (x_i - x_j),$$

onde:

$D_{(ij)}^2$ = distância entre i e j ;

x_i = vetor de observações da população i ;

x_j = vetor de observações da população j ;

S^{-1} = inversa da matriz comum de covariâncias para as n

populações.

Quando a distância euclidiana é calculada a partir de escores* (que estão livres dos efeitos de correlação entre variáveis), seu valor é análogo à distância de Mahalanobis. Sendo o cálculo de distância euclidiana mais rápido, a utilização desses escores propicia um importante ganho no tempo de computação.

A distância de Mahalanobis pode se aplicar a caracteres biométricos ou sobre marcadores moleculares (LEFORT-BUSON, 1985), entretanto, a sua aplicação com dados moleculares tem sido muito limitada devido a natureza não paramétrica de grande parte destes dados.

2.4. Distâncias entre populações caracterizadas por atributos qualitativos

Existem estimadores de distância, quando os atributos avaliados nos genótipos são do tipo qualitativo. Estas distâncias são muito usadas com dados de presença ou ausência de bandas eletroforéticas, seja estas de isoenzimas ou de proteína total. Numerosos são os índices de similaridade (assim chamados de estimadores de distância) propostos. Assim tem-se: de Jaccard (1908), de Julzynski (1927), de Dice (1927) e de Sokal & Sneath (1963), todos citados por

* Escore: Resultado obtido através da função discriminante para cada uma das variedades. Neste valor resume-se o conjunto de variáveis que caracterizam um genótipo, as quais têm a vantagem de não serem correlacionadas.

JACQUARD (1974). Outras distâncias utilizadas são as de Gower e a de Roger citadas por WRIGHT (1978).

Esta produção de índices externa a dificuldade de encontrar uma medição objetiva de semelhança (JACQUARD, 1974). É óbvio que se concluirá, algumas vezes, que o objeto *i* é mais similar ao objeto *k* que ao objeto *l*, e algumas vezes o oposto, dependendo de qual foi o índice usado.

Sendo o conceito de distância genética a soma-tória das alterações ao nível gênico, e portanto mais precisa quanto maior o número de genes avaliados, o uso de estimadores de distância quantitativos que englobam tanto as variações estruturais quanto regulatórias, permite teoricamente maior precisão de que as análises qualitativas (DAMERVAL *et alii*, 1987).

2.5. Relações genéticas a partir da avaliação quantitativa de bandas eletroforéticas de proteínas de semente

O extrato protéico bruto, de semente ou de outro tecido, quando fracionada num meio gel conveniente, produz um espectro de bandas que é um diagnóstico da espécie ou população. A homologia de bandas de diferentes espécies (ou populações) tem fornecido um critério de afinidade genética, que tem auxiliado aos evolucionistas nas suas inferências (LADIZINSKY & HYMOWITZ, 1979). No entanto, o coeficiente de semelhança, utilizado por estes pesquisadores, na maioria dos casos, é o índice de similaridade (% de simila-

ridade = [(número de pares de bandas similares) x 100]/(número de bandas diferentes + número de pares de bandas similares). Aos níveis taxonômicos de suas comparações (entre gêneros e espécies) são encontradas diferenças que permitem fazer inferências das relações filogenéticas. Ainda ao nível de caracterização intraespecífica, esse índice de similaridade também foi eficiente (SIDDIQ *et alii*, 1972, em arroz; BUSHUK & ZILLMAN, 1978, em trigo; MARCHYLO & LAMBERGE, 1980, em cevada). Entretanto, comparações de variedades na base de suas diferenças quantitativas de frações protéicas têm sido levadas em conta. Uma evolução rápida dos avanços, nesta abordagem, é feita a seguir.

Um dos primeiros usos de perfis densitométricos, em eletroforese de proteínas, está relacionado à identificação de bandas homólogas entre espectros com velocidades de migração discrepantes (JOHNSON *et alii*, 1967). O método requer a eletroforese de uma mistura 1:1 de proteína de cada espécie com a proteína padrão, para facilitar a identificação de bandas homólogas.

Quanto a técnicas que permitam caracterizar variedades individuais, McDANIEL (1971) comparou diferentes cultivares, dentro de três espécies do gênero *Hordeum* e *Secale cereale*. As proteínas de diferentes cultivares foram comparadas com base na similaridade qualitativa e quantitativa. O número de bandas proteicas comuns de similar concentração foi estimado usando-se um densitômetro. Bandas que diferiram menos que 25% nas medidas densitométricas da in-

tensidade foram consideradas como as mesmas, quantitativamente. No final, o número de bandas proteicas que dois cultivares tinham em comum foi expressado como uma porcentagem do número total de bandas visualizadas. McDANIEL (1971) achou que cultivares relacionados por descendência foram mais similares no fenótipo das proteínas do que cultivares não aparentados. Também para confirmar as diferenças qualitativas em bandas eletroforéticas de proteínas, medições densitográficas foram efetuadas por UNCAR & BOUCARD (1974) no gênero *Suaeda* e por PANDA *et alii* (1986), em *Capsicum*.

Um dos primeiros trabalhos que utilizam intensidades de bandas, nas fórmulas para identificar cultivares, foi o de ZILMAN & BUSHUK (1979). A citada quantificação consistiu em uma escala subjetiva de 1 a 5, baseada na projeção do negativo fotográfico dos espectros eletroforéticos. Eles acharam fórmulas de eletroforegramas para as 18 variedades de trigo estudadas. Materiais estreitamente relacionados por pedigree, no geral, tinham ao menos uma banda, na sua fórmula, que o distinguia do seu parente próximo. Igualmente, trigos duros foram facilmente distinguidos dos trigos comuns.

Uma abordagem estatística um pouco mais acurada da variação quantitativa de bandas eletroforéticas de proteínas obtidas de SDS-PAGE foi feita em arroz. SARKAR & BOSE (1984) calcularam as áreas que correspondem aos picos de leitura densitométrica das bandas e consideraram como porcentagem da área total de todos os picos. A análise de

variância dos picos de globulinas de mobilidades relativas (R_f) de 0,27 e 0,43 mostrou diferenças significativas. Os efeitos ambientais devidos a localidades e anos não foram significativos. Os mesmos autores concluíram que a distinção dos perfis eletroforéticos de cada variedade é, essencialmente, sob controle de fatores genéticos, e que esta variação pode ser convenientemente utilizada na identificação de cultivares. DERBYSHIRE *et alii* (1981) compartilham a idéia de SARKAR & BOSE (1984) e afirma que os perfis obtidos de géis de poliacrilamida estão relacionados á afinidade genética das amostras.

Um trabalho mais completo foi realizado por ALIAGA-MOREL *et alii* (1987) com 43 variedades de arroz, estudando a obtenção de medidas mais objetivas, estáveis e reproduzíveis para a identificação de variedades. A estimativa da afinidade entre as variedades sobre "bases mais seguras", segundo a expressão dos autores, foi o objetivo complementar. Um elevado número de caracteres morfológicos qualitativo e quantitativos, assim como medidas densitométricas (57 bandas) de perfis eletroforéticos foi registrado. A similaridade entre variáveis foi estimada a partir de um coeficiente de distância em enfoque multivariado. Dos resultados obtidos ALIAGA-MOREL *et alii* (1987) concluíram que a análise quantitativa de componentes (frações) de proteínas de embrião fornece maiores conhecimentos de afinidade entre variedades, do que aqueles obtidos a partir de caracteres morfológicos.

A habilidade dos dados bioquímicos (quando avaliados quantitativamente) de proporcionar descrições altamente discriminatórias foi salientada também por SMITH & SMITH (1989). Eles compararam 34 dados morfológicos com herdabilidade maior que 50%, dados iso-enzimáticos de 31 loci e dados cromatográficos de zeína em 31 linhas endogâmicas de milho. Depois de correlacionar dados de heterose com as distâncias calculadas com os diferentes tipos de dados mencionados, eles concluíram que os dados morfológicos não podem ser confiavelmente usados para estimar distâncias entre cultivares de milho que refletem as suas associações genéticas ou de "pedigree". Por sua vez, os dados bioquímicos poderão ser usados na descrição de linhas e híbridos de milho. Finalmente, os mesmos autores ressaltaram que, quando as intensidades de bandas foram consideradas nos dados cromatográficos, a sensibilidade foi acrescentada na distinção de linhas.

2.6. Análise de conglomeração

As relações grupais entre pares de indivíduos, cujas distâncias têm sido calculadas, podem ser melhor apreciadas por meio de agrupamentos. A análise de conglomeração visa agrupar os objetos, populações, etc., em classes, de tal forma que similares estejam no mesmo grupo.

Muitos são os algoritmos propostos para a análise de conglomeração. Entretanto a atenção estará restrin-

gida a duas abordagens particulares muito usadas em melhoramento.

O método Tocher começa com duas entidades estreitamente associadas e encontra uma terceira entidade que tem a menor média de D^2 junto com os dois primeiros ($\Sigma D^2/n$; n = número de distâncias envolvidas = 3 neste caso). Similarmente o quarto é escolhido por ter a menor média de D^2 junto com os três anteriores (a média é $\Sigma D^2/n = 6$), e assim sucessivamente. Se em qualquer estágio a média do D^2 de um grupo de aqueles que já tem sido considerados aparece alto, então a última entidade adicionada é retirada, passando a constituir-se um grupo com os componentes anteriormente examinados (RAO, 1952).

No método do vizinho mais próximo, todos os indivíduos começam sós num grupo de um e são posteriormente fusionados em grupos sobre um dado nível de distância. Esse nível de distância vai crescendo com as grandezas das distâncias até atingir a todos os indivíduos num só grupo (MANLY, 1986). O método de Tocher fornece uma relação de agrupamentos, entretanto o método do vizinho mais próximo dá uma relação hierárquica de afinidade entre genótipos, o que torna estes dois métodos em parte complementares. Uma revisão compreensiva sobre agrupamentos pode ser vista em RAO (1952), SINGH & CHAUDARY (1977) e MANLY (1986).

Extensa é a literatura das aplicações da análise de conglomerção no melhoramento, CRUZ (1990) fornece numerosas referências à respeito. O uso das análises de

agrupamentos hierárquicos (aglomerativos) como ferramenta no manejo de variação em coleções de germoplasma é destacada por PEETERS & MARTINELLI (1989). Os resultados da análise de agrupamento, segundo eles, poderão ser usados para classificar entradas de coleções de germoplasma, de acordo com seus respectivos grupos ("pool") gênicos, quando nenhum dado de passaporte* for disponível. Ainda concluem que, baseada nesta técnica, o potencial segregativos das entradas pode ser estimado.

2.7. Análise de função discriminante

Através da análise de função discriminante (AFD), é possível separar dois ou mais grupos de indivíduos, a partir de medidas desses indivíduos em várias variáveis (MANLY, 1986). Por exemplo, HUSSAINI *et alii* (1977) avaliaram 18 caracteres em 640 estoques genéticos de *Eleusine coreana* e estudaram a diversidade desses materiais aplicando a AFD, além de uma outra técnica utilizada neste tipo de trabalho, chamada análise de componentes principais (ACP). A AFD distinguiu os materiais em grupos razoavelmente semelhantes aos grupos formados pela ACP, com exceção de um, que foi subdividido em dois, os hindus e os etíopes. A AFD, portanto, demonstrou ser a mais sensível que a ACP.

* Os dados de passaporte são aqueles que fornecem fundamentalmente informação da procedência das entradas de um banco de germoplasma.

MORISHIMA & OKA (1981) também testaram o poder discriminatório de 11 caracteres na diferenciação dos subgrupos Índica e Japônica em 89 cultivares de uma coleção asiática de arroz. As funções discriminantes combinando valores de três ou quatro caracteres, possibilitaram a classificação incorreta das linhagens e subgrupos com um máximo de 0,7% de erro. No entanto, foram percebidos no gráfico de dispersão alguns tipos intermediários com características combinadas dos grupos Índica e Japônica. Contudo, uma aplicação mais interessante da AFD foi feita por OKA & CHANG (1962). Eles juntaram, em uma avaliação, diversos materiais de *Oryza perennis* ("spontanea", intermediária e "perennis"), linhas de Jeipore¹ (selvagem e cultivado) e *Oryza sativa* (tipo Índica, Japônica). Com a finalidade de avaliar as variações como um todo (ao nível multivariado), fórmulas discriminantes para classificar tipos perenes e espontâneos, formas selvagens e cultivadas e os tipos Índica e Japônica foram estabelecidas para cada tipo, usando dados de linhas selvagens e cultivadas como controle. Examinando os escores resultantes dessas funções discriminantes. Os autores observaram que as linhas de Jeipore foram em maior número intermediárias entre selvagens e cultivadas e que aquelas formas selvagens eram do tipo "perennis". Isto sugere, segundo eles, que o tipo "perennis" do arroz cultivado poderia ser o

¹ Região montanhosa da Índia, onde supõem os autores, encontram-se plantas intermediárias da evolução do arroz, de formas selvagens a espécies cultivadas.

seu progenitor.

As conclusões que podem ser tiradas da AFD derivam, muitas vezes, de apreciações visuais de diagramas de dispersão dos primeiros componentes, quando estes acumulam suficiente variabilidade (geralmente mais que 70%). Assim, os dois primeiros componentes da AFD de ERSKINE *et alii* (1989) acumularam 90% da variabilidade total sobre uma análise por país de origem entre os 13 primeiros produtores mundiais de lentilha. O agrupamento por país de origem foi sumarizado visualmente em um gráfico de dispersão de duas dimensões. Foi clara a formação do grupo constituído por acessos do Egito, Jordânia, Líbano e Síria. Um segundo grupo compreendeu acessos da Grécia, Irã, Turquia e Rússia.

Mas, nem todos estão satisfeitos com o uso da AFD. RISI & GALWEY (1989) opinaram que, para o seu trabalho, tanto a AFD como a ACP não forneceram informações claras. A AFD distinguiu entre ecótipos de *Chenopodium quinoa* Willd, mas a classificação obtida foi qualificada de arbitrária, exceto no caso de ecótipos subtropicais.

Embora tenha outras aplicações, até mais úteis, a AFD está sendo mais usado em uma de suas aplicações no melhoramento de plantas. Geralmente, serve para confirmar as relações de semelhança reportadas com a ACP (COSME, 1990; VARMA & GULATI, 1982), e a distância de Mahalanobis (DAS & GUPTA, 1984; RAO *et alii*, 1981; JAIN *et alii*, 1981; PETER & RAI, 1976; RAM & PANWAR, 1970).

3. MATERIAL E MÉTODOS

3.1. Materiais

Foram tomados ao acaso 67 genótipos de arroz da coleção de germoplasmas do Instituto de Genética, da Escola Superior de Agricultura "Luiz de Queiroz"/USP, Piracicaba - SP. Desses 67 germoplasmas, 58 são "brasileiros", provenientes do CNPAF/EMBRAPA, Goiânia, Goiás, e do IAC, Campinas, SP e 9 são japoneses provenientes da Estação Experimental de Arroz de Sequeiro, Japão. A relação dos germoplasmas utilizados é apresentada na Tabela 1.

Tabela 1 - Germoplasmas brasileiros e japoneses utilizados no presente trabalho.

NÚMERO	NOME	PROCEDÊNCIA	BANCO DE ORIGEM	TIPO
1	Bico Ganga	GO ¹	CNPAF	M
2	Guaíra	-	CNPAF	O
3	Dourado Precoce	SP	CNPAF	M
4	4 meses	-	CNPAF	O
5	Guedes	-	CNPAF	O
6	Amarelão*	MG ²	CNPAF	M
7	Pratão Precoce	SP	CNPAF	M
8	Pratão	-	CNPAF	M
9	Matão	-	CNPAF	O
10	Fernandez	GO ¹	CNPAF	O

continua.

Tabela 1 - Continuação.

NÚMERO	NOME	PROCEDÊNCIA	BANCO DE ORIGEM	TIPO
11	IAC-5544	SP	CNPAF	M
12	IAC-25	SP	CNPAF	M
13	IAC-1246	SP	CNPAF	M
14	IAC-120*	SP	CNPAF	M
15	IAC-47	SP	CNPAF	M
16	IAC-5032	SP	CNPAF	M
17	Iguapão	-	CNPAF	O
18	Iguape	-	CNPAF	O
19	IAC-416*	SP	CNPAF	M
20	Montanha Liso	-	CNPAF	O
21	Pérola de Campinas	-	CNPAF	O
22	Bacaba	-	CNPAF	O
23	Procópio	-	CNPAF	O
24	Caqui*	PB ²	CNPAF	O
25	Agulha Branca*	PB ²	CNPAF	O
26	Ourinho	-	CNPAF	O
27	Bico Preto	-	CNPAF	O
28	Aguihão	-	CNPAF	O
29	Batatais	-	CNPAF	M
30	Zebú	MA ²	CNPAF	O
31	Agulha ESAV	MG	CNPAF	M
32	Patnaí 6	-	IAC	M
33	Dourado Peludo	-	IAC	O
34	Dourado Liso	-	IAC	O
35	Dourado Agulha	SP	IAC	O
36	Paraíba	-	IAC	O
37	Cateto XXII	-	IAC	M
38	Cateto Dourado	-	IAC	M
39	Catetão	-	IAC	O
40	Dourado	-	IAC	M
41	Dourado Seco	-	IAC	M
42	Matão Branco	-	IAC	O
43	Pratão de Campinas	-	IAC	O
44	Goiano	-	IAC	O
45	Pindorama	-	IAC	O
46	4 meses N ^o XL	-	IAC	M
47	Dourado Agulha 2	-	IAC	M
48	Birigui	-	IAC	O
49	Itarabé	-	IAC	O
50	Santa América	SP ²	IAC	O

continua.

Tabela 1 - Continuação.

NÚMERO	NOME	PROCEDENCIA	BANCO DE ORIGEM	TIPO
51	Pratão IAMG Nº 11	MG	IAC	M
52	Pratão Uberlândia	-	IAC	O
53	Amarelão IAMG Nº 2	MG	IAC	M
54	Piemonte IAMG Nº 4	MG	IAC	M
55	Amarelão IAMG Nº 14	MG	IAC	M
56	Jaguari ESAV IAMG Nº 25	MG	IAC	M
57	Batataes	-	IAC	M
58	IAC-1391	SP	IAC	M
59	Dango Mochi	-	Japão	M
60	Kahee	-	Japão	M
61	Shinkoku Daio Mochi	-	Japão	M
62	Norin Mochi 4	-	Japão	M
63	Kazusa Wase	-	Japão	M
64	Senshōw	-	Japão	M
65	Oohata Wase	-	Japão	M
66	Gaisen Mochi	-	Japão	M
67	Gose Yonkoku	-	Japão	M

¹ Segundo FONSECA & BEDENDO (1984).

² Segundo FONSECA et alii (1981).

* Cultivares de arroz irrigado.

CNPAF = Centro Nacional de Pesquisa de Arroz e Feijão/EMBRAPA.

IAC = Instituto Agrônomo de Campinas.

O tipo M indica germoplasmas com algum melhoramento por instituições; O significa "não melhorado" e envolve variedades originais ou antigas com baixo grau de melhoramento. A informação para esta classificação veio do IAC e CNPAF, além das bibliografias citadas. Entretanto, em alguns poucos casos, segundo especialistas a delimitação entre M e N não é clara.

3.2. Métodos

3.2.1. Trabalhos no campo

Todos os germoplasmas utilizados foram cultivados sob idênticas condições, dentro do plano de renovação e manutenção da coleção de arroz do Departamento de Genética, ESALQ/USP. Foram semeadas em 20.10.88 no campo experimental da AGROCERES S.A., em Santa Cruz das Palmeiras - SP, e a colheita foi até 04.04.89 com materiais mais tardios, sendo os germoplasmas japoneses bem mais precoces. A semeadura foi realizada em sulcos corridos, em fileiras de 2,0 m por germoplasma, sem repetição, com uma quantidade média de 75 sementes por metro linear. A colheita foi manual e foram seguidas as práticas usuais de cultivo.

3.2.2. Trabalhos no laboratório

3.2.2.1. Extração da proteína solúvel

As proteínas foram extraídas de 15 grãos secos (14% de umidade e 10 meses de armazenamento em condições de ambiente), descascados e triturados em almofariz. Para cada germoplasma, 50 mg de pó foi homogeneizado com 200 ml de tampão de extração*, e logo a seguir agitado por 30 segundos, sendo deixado em repouso durante duas horas e logo centrifugado por dois minutos a 15.000 rpm.

* As composições desta e outras se encontram na Tabela 3.

3.2.2.2. Quantificação de proteína

A solução de reagente de Bradford (Tabela 2), diluída em água na proporção de 1:2, foi distribuída em cinco tubos de ensaio. Nestes tubos, foram acrescentadas frações de solução de albumina bovina (Tabela 2) correspondentes a 0, 10, 20, 50 e 100 μg de proteína para a construção de curva padrão. As amostras foram separadas adicionando-se 10 ml de proteína solúvel em outros cinco com 5,0 ml da solução reagente de Bradford. Todos os dez tubos foram homogeneizados em vortex. Os valores de absorção foram obtidos em espectrofotômetro a 595 nm. Com os valores dos padrões, foi construída uma curva, utilizada para o cálculo das concentrações proteicas das amostras.

Tabela 2 - Composição de soluções para quantificação de proteínas pelo método de BRADFORD (1976).

Reagente de Bradford:	100 mg do corante Coomassie - brilliant blue G-250, diluído em 50 ml de etanol, e misturado em 100 ml de ácido fosfórico 85%.
Solução estoque de albumina bovina	0,1 g de soro albumina bovina em 100 ml de água.

Tabela 3 - Composição das soluções estoques do procedimento eletroforético SDS-PAGE.

-
1. Tampão de extração: Solução de trisma 10 mM, com pH ajustado a 8,0 com HCl e 0,5% de 2 mercapto-etanol.
 2. Solução de acrilamida/bis-acrilamida (A/B): acrilamida 30 bis-acrilamida 0,8 g
 3. Tampão do gel inferior: Solução trisma 1,5 M, com pH ajustado a 8,8, com HCl.
 4. Tampão do gel superior: Solução trisma 0,5 M ajustado o valor do pH para 6,7 com HCl.
 5. Solução SDS (sódio duodecil sulfato): 1,0 g de SDS, ajustado para 10 ml de água (10%).
 6. Solução de persulfato de amônio (PSA) 10%: Essa solução não deve ser estocada, devendo-se prepará-la cada eletroforese (no caso se preparava cada vez 0,1 g/1 ml).
 7. Tampão de corrida (estoque): Trisma 12,0 g
Glicina 57,6 g
Ajustar para 1 litro com água destilada (pH da solução ~ 8,3)
 8. Tampão de amostra: 1,25 ml do tampão do gel inferior
2,0 ml de glicerol
Completar para 6,0 ml com água destilada e deionizada.
 9. Tampão de amostra desnaturante:
500 µl de 2 mercaptoetanol
250 µl de azul de bromofenol 0,1%
4 ml SDS (10%)
5,3 ml de tampão de amostra
-

3.2.2.3. Preparação de amostra denaturada

As amostras de proteína solúvel são ajustadas com tampão de extração (Tabela 3) para obtenção de concentração final de 70 µg/50 µl de solução protéica. Prepararam-se as amostras, colocando-se 50 µl da solução de proteí-

na solúvel para 50 μ l de tampão de amostra desnaturante. As misturas foram submetidas a temperatura de 100°C por cinco minutos, previamente à eletroforese.

3.2.2.4. Eletroforese SDS-PAGE

As proteínas foram separadas eletroforeticamente em um sistema bifásico de gel de poliacrilamida, segundo a metodologia de LAEMMLI (1970). Este sistema utiliza um gel superior (chamado espaçador) com baixa concentração de acrilamida (4,5%) e pH 6,7 com a função de concentrar as proteínas, e, um gel inferior (separador) com alta concentração de acrilamida (12%) e pH 8,9 no qual os polipeptídeos são separados de acordo com seu peso molecular.

Os géis utilizados foram de 10 x 13 cm e de 1 mm de espessura, sendo 7,5 cm de gel separador e 2,5 cm de gel espaçador. Para a formação dos géis, as placas foram seladas com solução de agar 1% e o preparo dos mesmos foi a partir de soluções estoques descritas na Tabela 3. A composição e modo de preparação dos géis estão apresentados na Tabela 4.

Tabela 4 - Composição dos géis (SDS-PAGE) (roteiro de preparação).

COMPONENTES	GEL SEPARADOR (12% ACRILAMIDA)	GEL ESPAÇADOR (4,5% ACRILAMIDA)
Tampão superior (4) *	-	2,5 ml
Tampão inferior (3)	10,0 ml	-
Solução acrilamida-bis (2)	16,0 ml	1,5 ml
Água deionizada	12,5 ml	5,4 ml
Solução SDS 10% (5)	0,4 ml	0,1 ml
Solução PSA 10% (6)	0,4 ml	0,1 ml
Temed	10,0 μ l	10,0 μ l

* Os números entre parênteses correspondem à numeração de solução da Tabela 2.

Para a utilização do tampão de corrida (solução 7 da Tabela 3), o tampão estoque foi diluído em água na proporção de 1:4 e ajustado para 0,1% de SDS.

Para a corrida eletroforética, 40 μ l de extrato proteico foi aplicado em cada canaleta do gel, e submetido a um ambiente a 4°C de temperatura, com intensidades crescentes de voltagem conforme o seguinte esquema: 50 volts (V), desde o início da corrida até as amostras saírem das canaletas; 70 V no percurso do gel espaçador; 150 V até 2/3 do gel separador; 170 V até a terminação do gel.

Ao final da corrida, os géis foram corados e fixados com uma solução mista de metanol: ácido acético, e água (30:7:63) e Coomassie-brilliant-blue (0,05%) durante uma noite. O descorado foi feito com várias trocas da solução metanol: ácido acético: água (30:7:63).

Para estimativa dos pesos moleculares aproximados das bandas eletroforéticas, foi aplicada em cada gel uma amostra de pesos moleculares padrões (marca Pharmacia)

com polipetídeos de 94, 67, 43, 30, 20,1 e 14,4 kD, que corriam junto às amostras.

3.2.2.5. Densitometria

As características das frações proteicas foram medidas por um densitômetro Tecnow, resultando em perfis densitométricos de cada variedade. A medida aproximada da concentração de cada fração foi obtida considerando-se a altura dos picos correspondentes. Para a utilização posterior, os valores dos picos foram corrigidos proporcionalmente, para que ficassem referidos a uma quantidade total padrão de proteína analisada. A seguinte fórmula de correção foi utilizada:

$$VC_{ij} = VO_{ij} + (QM - VO_{ij})/NP$$

onde:

- VC_{ij} = Valor corrigido da repetição i do cultivar j;
- VO_{ij} = Valor observado da repetição i do cultivar j;
- QM = Quantidade total padrão de proteína analisada (medida pela somatória máxima dos valores dos picos, e que correspondeu a 454,5 na variedade Oohata Wase);
- NP = Número de picos analisados (frações).

Os dados densitométricos corrigidos não entraram diretamente no cálculo das distâncias, mas previamente foram obtidos escores discriminantes (de 8 componentes e 97,2% de variância) a partir deles, por razões expostas no item 3.3.2.2.

3.3. Metodologia Estatísticas

3.3.1. Análise univariada

Visando estabelecer diferenças significativas nas respostas médias das variedades, foi realizada uma análise de variância. Esta análise permitiu determinar as variáveis com maior poder de diferenciação. Devido à forma como foi conduzido o experimento, onde as avaliações em géis, embora uniformes, foram preparadas em dias diferentes, um delineamento em blocos ao acaso com duas repetições foi considerado. Este delineamento permite distribuir o erro no contexto de todo o experimento.

O modelo matemático do delineamento em blocos ao acaso é:

$$Y_{ij} = m + t_i + b_j + e_{ij}$$

onde:

- Y_{ij} = observação do tratamento i ($i = 1, 2, \dots, v$) na repetição j ($j = 1, 2, \dots, r$);
- m = média geral do experimento;
- t_i = efeito de tratamentos i ;
- b_j = efeito do bloco j ;
- e_{ij} = erro experimental associado ao tratamento i na repetição j .

Para se detectar se existem diferenças entre variedades dentro do modelo, usou-se as seguintes hipóteses nula (H_0) e alternativa (H_a):

$$H_0 = m_1 = m_2 = \dots = m_{58} \quad \text{vs} \quad H_a = m_i \neq m_j$$

para todo $i, j = 1, 2, \dots, 58$

Nestas hipóteses, procura-se encontrar ao menos duas variedades que respondam diferente à resposta média. Para esse fim foi aplicado o teste de F ($F = QMV/QMR$), que se distribui com graus de liberdade da variedade (glV), graus de liberdade do resíduo (glR) e um nível de significância α .

onde:

QMV = quadrado médio de variedades

QMR = quadrado médio do resíduo

O teste de F foi feito a partir da análise de variância (Tabela 5).

Tabela 5 - Esquema da análise da variância em blocos ao acaso.

FONTE DE VARIAÇÃO	GL	SQ	QM	F
Blocos	$r-1$			
Variedades	$v-1$	SQV	QMV	QMV/QMR
Resíduo	$(r-1)(v-1)$	SQR	QMR	
TOTAL	$rv-1$			

3.3.2. Análise multivariada

No contexto da análise multivariada dos dados, visando encontrar as variáveis que mais caracterizavam os grupos e, portanto, as variáveis que apresentavam maior semelhança, montou-se a matriz de distâncias:

$$\Delta = i \begin{matrix} & i' \\ \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1p} \\ d_{21} & d_{22} & \dots & d_{2p} \\ \vdots & \vdots & d_{ii'} & \vdots \\ d_{p1} & d_{p2} & \dots & d_{pp} \end{bmatrix} \end{matrix}$$

onde, $d_{ii'}$ representam a distância euclidiana média entre os pares de variáveis i , i' , a qual é obtida a partir da expressão:

$$d_{ii'} = \frac{1}{\sqrt{p}} \left[\sum_{j=1}^p (x_{ij} - x_{i'j})^2 \right]^{1/2}$$

onde:

p = número de características (variáveis) estudadas.

3.3.2.1. Agrupamento das variedades

Uma vez obtida a matriz de distâncias, procedeu-se o agrupamento de variedades semelhantes usando-se alguns algoritmos conhecidos no processo de agrupamento. Os métodos de agrupamento podem ser aglomerativos ou divisivos.

Os métodos aglomerativos partem do fato de que cada observação forma um grupo e, em etapas sucessivas, vai formando agrupamentos semelhantes até chegar a formar um só grupo. O método aglomerativo do vizinho mais próximo ("Single Linkage Method") foi usado na formação dos grupos. Para se entender a formação operativa deste método, supõem-se que haja:

Grupos

1 2 3

Número de observação por grupo

n_1 n_2 n_3

e, (1, 2) simbolizam 1 e 2 unidos

$$n_{1,2} = n_1 + n_2$$

A medida de similaridade do grupo é obtida a partir da distância:

$$d_{(1,2),3} = \alpha_1 d_{1,3} + \alpha_2 d_{2,3} + \beta d_{1,2} + \gamma [d_{1,3} - d_{2,3}]$$

Esta é uma expressão geral de agrupamento obtida por Lance & Williams, citado por CURI (1983);

onde:

$$\alpha_1 = \alpha_2 = 1/2, \quad \beta = 0 \quad \gamma = 1/2$$

$$d_{(1,2),3} = \frac{1}{2} d_{1,3} + \frac{1}{2} d_{2,3} + \frac{1}{2} [d_{1,3} - d_{2,3}]$$

onde, d_{ij} é obtido da matriz de distâncias euclidianas médias Δ .

O método divisivo que também foi usado no presente trabalho consiste em um procedimento contrário do agrupamento do caso anterior, ou seja, partindo de um só grupo onde se encontram todas as observações, vai se dividindo em subgrupos. O método divisivo de Tocher usado nesta tese, segue este procedimento. O par de genótipos mais próximos (d^2 menor) constituirá o primeiro grupo. A seguir, é analisada a possibilidade de entrada de novos genótipos ao grupo inicial, o que é aceito se o acréscimo na média de d^2 intragrupo, devido sua incorporação, for menor que a maior

das distâncias mínimas, que envolve cada um dos genótipos. Ao contrário do método anterior, o método divisivo de Tocher agrupa as variedades, constituindo-se aglomerados independentes.

3.3.2.2. Análise de função discriminante (AFD)

Através do AFD é possível separar dois ou mais grupos de observações (no caso, variedades) onde foram medidas várias variáveis. Em termos gerais, o AFD é uma combinação linear:

$$Y = LX = (\underline{u}_i - \underline{u}_j)' \Sigma^{-1} X$$

onde, $\underline{u}_i$ e $\underline{u}_j$ representam os vetores de médias populacionais associadas aos grupos i e j, respectivamente. Σ^{-1} é a inversa da matriz de variâncias e covariâncias, a qual se considera que é comum para todas as observações em geral, donde:

$$\underline{u}_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{ip} \end{bmatrix}$$

para p variáveis.

A matriz de variâncias e covariâncias fica:

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22}^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp}^2 \end{bmatrix}$$

Como em geral $\underline{\mu}$ e Σ são desconhecidos, deve-se estimar, tendo que, neste caso, a função discriminante estimada fica:

$$\hat{Y} = (\bar{X}_i - \bar{X}_j)' S^{-1} \underline{X}$$

onde,

$$\bar{X}_i = \begin{bmatrix} \bar{X}_{i1} \\ \bar{X}_{i2} \\ \vdots \\ \bar{X}_{ip} \end{bmatrix}$$

$$S = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{i1} & S_{i2} & S_{ij} & S_{ip} \\ \vdots & \vdots & \vdots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix}$$

$$\bar{X}_{i1} = \frac{1}{n_i} \sum_{j=1}^p X_{ij}; \text{ média amostral}$$

e,

$$S_{ij} = \frac{1}{n_1} \sum_{j=1}^p (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_j); \text{ variância amostral}$$

Uma vez obtidos os componentes (Z_i) da função discriminante (variáveis canônicas), um número destas que envolvam um mínimo de 80% da variação disponível, é escolhido para a estimação de escores, para cada população. Com base nesses escores, que estão livres de correlação, estimou-se a distância euclidiana média. Esta distância por trabalhar em cima desses escores, adquire as propriedades da distância de Mahalanobis de dar estimados de distância independentes dos efeitos da correlação entre variáveis (GODOY, 1990, comunicação pessoal).

Após a obtenção da função discriminante, pode-se classificar as variedades segundo o seguinte critério, no caso de dois grupos:

$$\text{Se } (\bar{X}_2 - \bar{X}_1)S^{-1} X \geq \frac{1}{2}(\bar{X}_2 - \bar{X}_1)S^{-1} (\bar{X}_1 + \bar{X}_2),$$

então a nova observação é classificada no grupo um, em caso contrário no grupo dois.

Para se testar a validade ou não da AFD são usados muitos testes, particularmente o teste de Hotelling citado por GODOY (1985) e definido como:

$$U^{(S)} = \text{Traz } (R^{-1}H)$$

onde,

Traz é o traço de uma matriz;

$R = Y'[I - X(X'X)^{-1}X']Y$ = soma do quadrado do resíduo;

$H = B'C[C'X'X)^{-1}C] C'B$ = soma do quadrado da hipótese;

$S = \min \{q, p\}$ = número de autovalores diferentes de zero;

$$m' = \frac{1}{2} ([p-q] - 1)$$

$$n' = \frac{1}{2} (ne - p - 1)$$

$q = \rho(C)$ = amplitude da matriz C

$ne = n - \rho(X)$

n = número total de observação

$$Y = \text{matriz de observações} = \begin{bmatrix} y_{11} & y_{21} & \dots & y_{1\sigma 1} \\ \vdots & \vdots & & \vdots \\ y_{1p} & y_{2p} & & y_{1\sigma p} \end{bmatrix}$$

Matriz de
delinea-
mento
multivariado

$$X = \begin{bmatrix} \mu & \alpha_1 & \dots & \alpha_{58} & P_1 & P_2 \\ 1 & 1 & \dots & 0 & 1 & 0 \\ 1 & 1 & \dots & 0 & 0 & 1 \\ 1 & 0 & \dots & 0 & 1 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 1 & 0 & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 0 & 1 & 0 \\ 1 & 1 & \dots & 0 & 0 & 1 \\ 1 & 0 & \dots & 0 & 1 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 1 & 0 & \dots & 1 & 0 & 1 \end{bmatrix}$$

de forma tal que

$$\frac{2(Sn' + 1)}{S(2m' + S + 1)} U^{(S)} \stackrel{H_0}{\sim} F(2m' + S + 1, 2(Sn' + 1))$$

sendo H_0 a hipótese nula.

Todos os cálculos estatísticos foram feitos em computador. A análise de variância, distâncias e agrupamento foram feitos no SAEG-2, "Sistema de Análise Estatística e Genética" desenvolvido pela Universidade Federal de Viçosa.

A análise de função discriminante foi realizada no programa SISMUL - "Sistema Multivariado" de autoria do professor Cássio Godoy (Departamento de Matemática e Estatística, ESALQ). O SAS - "Sistema Analitical Statistic" foi usado para confirmar os agrupamentos. Todo o processo de análise estatístico teve a gentil assessoria do citado Prof. Cássio R. Godoy.

4. RESULTADOS E DISCUSSÃO

Para facilitar a sua apresentação, este capítulo foi dividido em três itens principais:

- Análise eletroforética e univariada;
- Distância entre variedades e agrupamentos;
- Aplicação da análise de função discriminante.

4.1. Análises eletroforética e univariada

No presente trabalho, foram analisados os perfis eletroforéticos de proteínas de cariopses, em géis de poliacrilamida, em condições desnaturantes (SDS-PAGE) de um total de 67 variedades de arroz, das quais 58 foram brasileiras com diferentes graus de melhoramento e 9 foram japonesas.

Através de análise densitométrica dos perfis, foi possível a determinação da concentração relativa de um total de 17* componentes proteicos, exemplificados na Figura 1. Tais componentes, em muitos casos, foram formados por

* Pela correção proporcional feita nos dados no item 2.2.2.5, uma variável teve que ser eliminada, para possibilitar os cálculos da AFD, segundo sugerido pelo assessor estatístico Prof. Cassio Godoy.

Figura 1 - Um típico gel SDS-PAGE e traços densitométricos de proteínas de cariopse de arroz (variedade IAC-5544). A direção de migração foi de direita para esquerda. Os valores dos 16 componentes proteicos, tomados em mm foram depois padronizados a valores relativos de acordo com o item 3.2.2.5.

mais de uma banda, devido à impossibilidade de diferenciação por parte do densitômetro entre bandas muito próximas. A maior parte das variações evidenciadas entre as variedades analisadas foi do tipo quantitativo e não qualitativo, como mostram os perfis da Figura 2. Os valores de concentração corrigidos para cada uma das 16 frações proteicas, são apresentados na Tabela 5, na qual cada um dos componentes é caracterizado por seu peso molecular aproximado, determinado com base nos padrões de peso molecular utilizando o "kit" de calibração de eletroforese (baixo peso molecular) da Pharmacia.

Figura 2 - Gel característico com os padrões eletroforéticos de 12 variedades. A direção de migração é de cima para abaixo com o primeiro conjunto da direita, correspondente ao padrão de pesos moleculares (Pharmacia LW).

Tabela 5 - Padrões eletroforéticos (SDS-PAGE) de 67 variedades de arroz mostrado 16 frações proteicas (variáveis) com seus respectivos pesos moleculares aparentes (*) em kd e valores corrigidos de concentração em mm, medidas no densitômetro.

CODIGO GENOT.	VAR.1	VAR.2	VAR.3	VAR.4	VAR.5	VAR.6	VAR.7	VAR.8	VAR.9	VAR.10	VAR.11	VAR.12	VAR.13	VAR.14	VAR.15	VAR.16
	39.8*	36.1	31.2	24.5	28.0	59.6	54.7	50.8	47.0	43.9	39.6	35.2	32.4	29.1	27.2	24.4
1	28.60	24.10	30.10	22.60	24.60	32.10	39.85	25.85	28.85	29.35	30.85	31.10	19.35	22.10	25.10	20.85
2	31.54	22.79	33.79	23.79	24.79	33.29	40.54	26.29	27.04	27.54	30.79	30.79	19.54	20.54	24.79	19.04
3	24.91	21.41	30.16	25.16	25.66	32.16	38.16	25.66	28.91	29.66	31.66	30.91	20.41	20.66	25.41	21.41
4	29.26	23.01	29.51	22.26	23.51	32.01	41.76	26.51	27.26	28.01	32.26	31.26	21.26	23.76	26.76	17.76
5	32.24	25.74	32.99	22.49	22.49	32.99	43.74	27.24	29.24	29.99	30.49	29.99	17.49	20.24	24.74	16.74
6	29.78	24.03	31.28	23.78	23.03	31.53	39.78	26.28	27.53	27.28	30.78	33.28	19.78	21.28	27.03	19.53
7	25.75	23.00	29.25	21.25	25.50	33.25	38.75	25.50	31.00	28.75	32.50	30.25	20.25	23.00	27.50	19.50
8	29.10	23.10	31.85	24.35	24.85	32.35	37.60	24.10	27.85	27.35	29.60	30.10	20.35	22.60	27.60	21.35
9	22.00	19.50	26.00	19.25	23.75	33.00	40.75	24.50	28.75	26.50	31.00	35.00	22.75	23.50	27.00	23.25
10	27.91	23.66	32.91	22.91	25.41	33.16	43.66	25.91	27.41	27.41	30.16	30.66	19.91	20.66	25.91	18.16
11	28.87	21.87	28.62	24.62	24.12	32.37	44.62	26.62	27.62	28.62	32.12	31.62	20.37	20.37	26.62	17.87
12	26.91	21.91	26.41	23.41	24.16	33.66	46.16	25.41	25.66	25.66	29.41	30.41	22.66	22.91	28.66	21.41
13	26.06	21.56	25.81	22.06	23.31	30.81	44.56	25.56	27.31	26.81	30.31	30.56	23.56	24.81	29.81	20.31
14	32.62	25.87	28.62	24.37	25.62	29.87	43.37	24.62	26.62	25.87	28.62	29.62	19.87	21.37	26.37	20.37
15	30.25	24.25	29.75	22.50	26.25	33.50	46.25	26.00	26.00	26.25	29.00	29.00	18.75	19.75	27.50	20.00
16	26.93	22.68	28.68	21.68	24.18	32.43	40.18	25.93	27.68	29.18	32.43	31.43	19.93	22.93	27.68	20.93
17	26.82	22.82	30.57	21.82	23.07	31.57	39.82	25.57	27.57	28.82	32.07	31.32	20.82	23.32	26.82	21.82
18	36.22	21.97	28.72	20.97	23.47	32.72	40.22	25.22	26.97	27.72	32.72	31.97	21.97	23.22	25.47	22.22
19	27.57	23.57	29.82	22.32	24.07	31.32	38.32	26.07	30.57	27.32	32.32	30.82	19.32	21.82	27.32	21.32
20	27.25	24.00	30.75	22.75	25.25	32.25	39.75	25.25	30.25	27.75	30.50	30.25	20.00	22.00	25.50	20.75
21	26.19	22.44	29.94	23.69	21.44	33.94	40.44	24.19	28.69	27.19	31.94	30.94	21.19	22.19	26.19	22.19
22	29.96	23.96	30.46	25.71	26.21	34.46	44.71	25.71	26.96	29.21	30.71	29.71	18.46	18.71	22.96	18.21
23	29.44	24.19	29.19	22.69	25.44	33.69	44.69	26.94	25.94	26.44	31.69	29.94	20.69	21.69	23.69	19.94
24	28.28	24.28	30.78	24.03	24.53	32.53	40.78	25.53	26.28	25.78	29.03	29.53	23.53	23.78	25.03	20.03
25	29.94	25.19	32.69	23.69	26.94	32.69	38.44	28.94	25.44	29.69	33.44	32.19	17.44	17.19	21.69	20.69
26	28.34	22.84	29.09	22.09	24.84	33.59	39.84	25.84	25.59	27.59	30.09	31.09	21.09	21.59	26.84	23.34
27	25.46	22.21	26.21	20.71	24.71	33.21	38.96	25.71	28.21	26.46	28.21	31.96	22.96	22.96	29.21	26.46
28	25.44	21.19	27.69	21.19	24.69	34.44	37.69	25.69	26.94	28.44	29.94	31.69	22.69	23.19	28.19	24.19
29	29.60	23.60	28.60	23.85	25.35	34.35	43.10	25.85	27.35	27.60	30.60	30.60	19.10	19.85	26.85	20.35
30	27.49	23.49	32.74	21.99	23.49	34.74	43.99	28.49	26.99	27.74	30.24	29.74	18.99	19.49	26.74	20.74
31	35.53	24.53	28.78	26.03	28.03	36.53	41.28	28.03	26.53	29.03	28.03	27.78	19.78	20.28	21.03	17.03
32	27.34	20.09	28.84	20.09	22.34	33.09	39.34	24.09	29.09	27.59	31.59	31.59	21.34	21.09	27.34	24.59
33	28.38	22.63	30.63	21.63	22.88	32.38	38.13	24.88	30.88	28.13	31.63	31.13	20.13	19.88	27.13	23.13
34	27.54	21.29	31.29	22.79	23.54	32.54	37.29	24.79	29.54	27.54	30.54	31.29	20.54	21.29	27.79	24.54
35	27.47	21.47	30.72	21.72	23.72	32.97	39.47	24.97	28.22	26.72	30.22	30.97	20.22	20.72	27.22	23.22

CONTINUA

Tabela 5 - Continuação.

CODIGO GENOT.	VAR.1	VAR.2	VAR.3	VAR.4	VAR.5	VAR.6	VAR.7	VAR.8	VAR.9	VAR.10	VAR.11	VAR.12	VAR.13	VAR.14	VAR.15	VAR.16
	89.8*	86.1	81.2	74.5	68.0	59.6	54.7	50.8	47.0	43.9	39.6	35.2	32.4	29.1	27.2	24.4
36	29.40	22.15	34.15	21.40	23.90	32.65	42.90	24.15	25.40	26.15	30.40	30.40	20.90	20.65	29.65	21.40
37	28.62	22.62	33.62	20.87	24.87	34.12	44.12	25.62	24.87	26.37	29.37	29.62	20.37	20.62	26.87	22.12
38	31.32	25.82	35.82	23.32	26.07	32.82	43.32	24.32	26.32	24.07	25.32	27.32	19.82	19.32	27.57	21.32
39	32.35	24.60	33.35	22.60	25.60	36.10	42.85	28.60	29.60	28.60	30.60	28.10	15.10	15.35	25.60	16.85
40	30.03	22.78	35.53	21.78	23.53	32.28	42.03	26.28	26.03	25.78	30.53	30.53	19.03	20.03	27.78	20.78
41	29.21	24.21	28.71	22.46	24.96	33.71	43.21	26.21	27.21	26.46	30.71	29.71	18.71	18.71	28.21	21.71
42	27.63	23.13	29.38	21.88	24.38	32.88	38.38	25.13	27.38	28.98	31.63	33.13	19.13	19.98	27.88	23.88
43	29.54	22.79	30.79	21.04	24.04	33.54	42.29	24.54	26.54	27.54	32.54	33.29	18.04	18.29	27.54	23.29
44	26.34	22.34	31.84	21.84	24.34	32.34	37.84	23.09	29.34	27.59	30.59	31.34	21.34	21.34	27.59	24.09
45	26.35	22.10	31.60	20.85	23.60	33.35	37.10	24.85	28.35	27.35	31.35	30.10	21.35	20.85	28.10	24.85
46	29.29	25.79	32.29	24.04	25.54	32.29	37.04	24.29	29.79	27.04	30.79	30.79	18.54	19.04	28.54	22.54
47	31.35	25.60	32.35	23.10	26.35	33.60	43.60	25.85	26.35	27.10	27.85	26.35	19.10	19.60	25.60	21.60
48	29.57	24.07	30.57	22.07	25.82	35.32	43.32	24.82	26.57	26.32	30.82	29.07	20.57	20.82	26.82	19.07
49	34.32	29.32	34.57	24.32	26.32	31.32	34.82	23.07	24.82	25.07	28.82	28.07	22.57	22.07	24.32	20.82
50	30.93	25.18	30.43	22.18	25.43	33.18	42.18	26.18	27.18	26.68	31.68	30.43	17.93	18.18	25.68	21.18
51	32.07	27.32	31.57	25.07	26.07	31.82	37.07	23.82	26.07	25.57	31.07	29.82	20.32	20.32	23.07	23.32
52	36.15	26.40	34.90	22.15	23.90	31.90	43.15	21.40	23.15	26.15	29.65	29.15	17.90	18.90	23.40	25.15
53	37.07	25.82	37.32	22.07	23.32	32.57	44.57	21.82	22.82	24.07	31.32	31.57	19.32	19.07	21.82	19.82
54	39.91	27.66	32.41	23.91	22.16	30.66	38.41	21.66	21.66	26.91	30.41	29.41	21.66	20.66	23.41	22.91
55	24.35	24.10	34.35	20.85	22.85	32.10	43.60	23.85	23.60	26.35	31.85	29.60	18.35	19.35	24.60	23.10
56	33.60	25.85	32.10	22.60	24.35	35.10	44.60	22.60	23.10	26.10	30.85	29.85	17.85	18.85	24.10	22.10
57	34.12	24.87	34.12	21.87	24.12	34.87	43.12	21.62	22.37	25.37	32.37	29.87	18.62	19.12	23.87	23.62
58	35.57	25.82	34.57	22.82	23.82	32.57	42.32	21.57	21.82	25.57	31.57	30.82	20.82	19.32	23.32	21.82
59	38.44	25.19	38.44	26.69	25.44	31.69	33.94	23.94	22.19	25.69	26.69	23.94	21.94	21.44	22.69	24.94
60	40.00	24.25	38.00	27.25	25.00	31.50	35.25	26.75	25.00	24.50	30.25	29.25	20.25	19.25	20.50	20.25
61	36.51	24.26	34.26	23.76	24.26	34.01	38.26	21.76	21.01	26.51	32.01	31.76	20.26	18.76	22.26	25.76
62	39.57	23.32	38.57	26.07	22.32	34.07	38.32	23.57	21.32	26.07	31.82	30.57	19.07	19.07	20.82	23.57
63	35.59	22.34	35.84	21.09	21.59	22.09	37.59	23.84	22.09	25.09	33.59	31.84	20.34	19.84	22.34	26.84
64	34.88	23.88	36.13	20.63	22.38	37.13	37.88	22.88	21.63	24.63	34.63	33.38	17.38	17.38	22.88	26.13
65	34.26	20.76	37.76	24.51	23.76	34.26	38.01	25.01	23.01	26.01	31.01	31.51	19.76	19.26	22.01	25.26
66	36.57	23.57	40.32	22.82	23.32	35.07	40.07	21.32	20.07	26.07	31.32	32.32	19.07	18.07	20.57	24.82
67	33.72	23.47	37.72	21.97	23.72	34.72	42.22	24.22	22.22	24.72	33.22	31.72	19.72	18.47	20.47	23.72

Visando examinar o comportamento da variabilidade de cada uma das variáveis, foi realizada uma análise univariada de variância. Os resultados da análise de variância de 58 variedades brasileiras, excluindo-se as japonesas*, foram resumidos na Tabela 6. Verifica-se que estas populações diferem significativamente, ao nível de 5% de probabilidade, em seis bandas eletroforéticas, mas não em dez delas.

* As variedades japonesas foram consideradas, neste estudo, fundamentalmente, para se fazer o contraste de origem no AFD.

Tabela 6 - Análise de variância de 16 caracteres protéicos em 58 variedades brasileiras no delineamento de blocos ao acaso. Piracicaba, 1989.

FONTES DE VARIACÃO	GL	QM DAS BANDAS				
		1	2	3	4	5
Blocos	1	186,62	109,44	124,44	1,12	62,45
Variedades	51	22,85*	6,97*	12,45*	3,84 ^{ns}	3,27*
Resíduo	57	9,69	3,51	5,61	2,74	1,69
Média		24,57	23,70	31,02	22,60	24,49
CVC(%)		10,52	7,91	7,63	7,32	5,31

FONTES DE VARIACÃO	GL	QM DAS BANDAS					
		6	7	8	9	10	11
Blocos	1	13,88	1,92	0,26	665,65	4,10	83,03
Variedades	51	3,11 ^{ns}	14,87	5,55*	9,55 ^{ns}	3,45 ^{ns}	4,19*
Resíduo	57	2,80	16,22	2,35	8,48	2,52	2,31
Média		33,01	41,17	23,70	26,95	27,19	30,66
CVC(%)		5,07	9,78	6,10	10,80	5,84	4,96

FONTES DE VARIACÃO	GL	QM DAS BANDAS				
		12	13	14	15	16
Blocos	1	99,87	5,09	134,35	1094,73	148,25
Variedades	51	4,33 ^{ns}	5,54 ^{ns}	6,56 ^{ns}	8,02 ^{ns}	9,41 ^{ns}
Resíduo	57	2,98	4,72	4,48	7,54	11,69
Média		30,48	20,04	20,77	26,15	21,39
CVC(%)		5,66	10,84	10,19	10,50	15,98

* : Significativo ao nível de 5% de probabilidade pelo teste F.

^{ns} : Não significativo.

Neste trabalho, como foi visto, a análise univariada de variância não foi efetiva na detecção da variabilidade, pois só seis de um total de 16 das variáveis

analisadas, apresentaram diferença significativas. Muito pelo contrário, a análise de variância multivariada deu diferença altamente significativa ao nível de menos de 1% para o efeito agregado de todos os caracteres, utilizando-se o teste de Hotelling ($n_1 = 912$; $n_2 = 802$; $F_{(obs)} = 1,41$). A esse respeito, DEMETRIO (1985) fez uma observação pertinente: em vários experimentos, onde são tomados dados de muitas variáveis, são feitas análises univariadas para cada uma dessas variáveis, individualmente, sem serem levadas em consideração as correlações existentes entre elas. São obtidas conclusões sobre o comportamento das variedades para cada uma das variáveis e, então, uma conclusão conjunta comum a todas elas é tirada, sem se levar em consideração um nível conjunto de significância. A análise de variância multivariada, ainda segundo DEMETRIO (1985), fornece resultados baseados na análise conjunta de todas as variáveis utilizadas. E, segundo WINER (1971), citado por DEMETRIO (1985), as respostas consideradas como uma única resposta, geralmente, contêm mais informações sobre o efeito total de tratamentos do que uma série de respostas consideradas separadamente.

Entretanto, ALIAGA-MOREL (1987) conseguiu diferenças significativas em análise univariada de variância em 37 frações proteicas, produto de eletroforese de proteínas de embrião de arroz. Nesta comparação, dois aspectos devem ser salientados: o densitômetro para a avaliação quantitativa de bandas usado por ALIAGA-MOREL (1987) é muito mais sensível ao utilizado no presente trabalho, e, isto é

confirmado pelo número de bandas detectadas por eles: o outro aspecto é o fato de ALIAGA-MOREL ter analisado o embrião. O tecido do embrião parece fornecer maior variabilidade que o cariopse (endosperma + embrião), por ser tecido $2n$. No caso do cariopse, existe muito mais proteína no endosperma ($3n$) que mascararia a variabilidade do embrião. Entretanto, não se pode, a princípio, concluir este item sem salientar que a extração manual do embrião de sementes de arroz seja uma tarefa muito trabalhosa e pouco prática, quando se deseja avaliar grande número de variedades.

Voltando a Tabela 6, do exame dos coeficientes de variação das variáveis (CV), pode-se observar que a precisão do experimento foi relativamente alta para a maioria das variáveis (10 delas) e intermediárias nas seis restantes.

4.2. Distâncias entre variedades e agrupamento

Os valores das distâncias euclidianas médias, para todas as possíveis combinações entre pares de variedades brasileiras são apresentados na Tabela 7. A magnitude das distâncias oscilou grandemente: a mínima foi 0,42 (entre Dourado Seco e Santa América) e a máxima 3,18 (entre Matão e Piemonte IAMG N. 4).

Tabela 7 - Continuação.

Cultivar	46	47	48	49	50	51	52	53	54	55	56	57	58
1. Bico Ganga	1.66	1.81	1.42	2.47	1.15	1.71	1.52	1.88	2.87	1.18	1.82	1.48	1.65
2. Guaira	1.25	1.43	1.36	2.36	1.28	1.65	1.77	1.43	2.88	1.38	1.22	1.39	1.25
3. Dourado Precoco	1.39	1.71	1.38	2.68	1.19	1.79	2.34	2.81	2.47	1.88	1.51	1.78	1.66
4.4. Neses	1.48	1.68	1.26	2.39	1.24	1.68	2.88	1.79	2.17	1.53	1.29	1.46	1.52
5. Suedes	2.88	1.51	1.86	2.62	1.64	2.82	1.66	1.84	1.91	1.28	1.38	1.58	1.74
6. Asarela	1.25	1.93	1.56	2.55	1.56	1.86	2.29	1.72	2.33	1.87	1.69	1.84	1.58
7. Pratao Precoco	1.47	1.42	0.77	2.83	0.91	1.26	2.11	1.89	2.48	1.58	1.17	1.26	1.63
8. Pratao	0.96	1.14	0.84	1.85	0.87	1.88	1.67	1.32	1.86	1.34	0.89	1.12	1.84
9. Atao	2.41	2.35	1.73	2.87	2.85	2.37	2.72	2.86	3.18	2.39	2.83	2.13	2.16
10. Fernandez	1.89	1.46	1.87	2.33	1.12	1.63	2.16	1.73	2.54	1.73	1.44	1.64	1.57
11. IAC 5544	1.56	2.18	1.77	3.89	1.57	2.13	2.51	2.24	2.55	2.88	1.81	2.85	1.85
12. IAC 25	1.61	1.87	1.42	2.69	1.54	1.86	2.58	2.28	2.77	2.33	1.73	2.89	1.94
13. IAC 1246	1.52	1.71	1.21	2.56	1.27	1.69	2.23	1.96	2.55	1.93	1.42	1.72	1.72
14. IAC 128	1.56	1.11	1.41	2.22	1.16	1.33	1.55	1.96	2.18	1.76	1.89	1.52	1.63
15. IAC 47	1.43	0.93	1.25	2.51	0.86	1.59	1.79	2.12	2.62	1.73	1.22	1.62	1.91
16. IAC 5832	1.59	1.46	1.35	2.65	1.86	1.76	1.96	2.82	2.41	1.39	1.25	1.47	1.78
17. Iguapao	1.59	1.58	1.44	2.59	1.26	1.84	1.98	1.74	2.27	1.28	1.38	1.45	1.68
18. Iguape	1.66	1.64	0.97	2.38	1.14	1.55	1.99	1.51	2.24	1.48	1.87	1.17	1.36
19. IAC 416	1.14	1.31	0.78	2.88	0.68	1.18	1.79	1.56	2.14	1.24	0.91	0.96	1.28
20. Montanha Lisa	1.28	0.94	0.71	1.98	0.88	1.17	1.67	1.49	2.11	1.34	0.83	1.12	1.32
21. Perola de Caspinas	1.76	2.17	1.43	2.39	1.71	1.88	2.43	1.54	2.13	1.89	1.68	1.65	1.38
22. Macaba	1.68	1.41	1.58	2.73	1.25	1.78	2.86	2.14	2.22	1.78	1.35	1.76	1.76
23. Procopio	1.48	1.15	0.82	2.21	0.83	1.38	1.92	1.98	2.31	1.61	0.99	1.38	1.62
24. Caqui	1.48	1.74	0.97	1.62	1.57	1.29	2.38	1.67	2.28	2.14	1.53	1.75	1.48
25. Agulha Branca	1.47	1.65	1.65	2.88	1.24	2.88	2.34	2.39	2.69	1.77	1.72	1.87	2.87
26. Gurinho	1.26	1.83	0.92	2.25	0.84	1.41	1.74	1.63	2.22	1.44	0.95	1.34	1.42
27. Bico Preto	1.77	1.53	1.48	2.46	1.62	1.98	2.21	2.82	2.82	2.22	1.68	2.84	1.99
28. Agulha	1.48	1.33	1.81	2.29	1.28	1.68	2.86	1.72	2.43	1.73	1.27	1.68	1.58
29. Batatais	1.15	1.26	1.11	2.47	0.88	1.45	1.98	1.95	2.27	1.74	1.17	1.58	1.57
30. Zebu	1.57	1.67	1.66	2.86	1.52	2.18	2.33	2.11	2.88	1.82	1.88	2.81	2.88
31. Agulha ESHV	2.51	1.55	2.16	2.63	2.11	2.89	2.18	2.67	2.18	2.36	1.81	2.32	2.38
32. Patnai 6	2.89	1.92	1.65	2.67	1.61	2.88	1.83	1.38	2.18	1.42	1.32	1.38	1.37
33. Dourado Peludo	1.46	1.66	1.34	2.33	1.26	1.61	1.77	1.14	1.98	1.38	1.15	1.15	0.98
34. Dourado Lisa	1.32	1.51	1.37	2.44	1.23	1.72	1.74	1.26	2.11	1.36	1.23	1.36	1.18
35. Dourado Agulha	1.57	1.48	1.14	2.31	1.15	1.63	1.76	1.25	2.25	1.35	1.18	1.17	1.25
36. Paraíba	1.88	1.89	1.38	2.51	1.38	1.82	2.21	1.53	2.63	1.72	1.65	1.68	1.46
37. Cateto XXI	1.46	1.26	1.32	2.48	1.17	1.84	1.88	1.63	2.62	1.46	1.36	1.53	1.68
38. Cateto dourado	1.52	1.46	1.55	1.92	1.67	1.69	2.86	1.79	2.68	2.17	1.77	1.94	1.84
39. Catetao	1.71	1.48	1.61	2.63	1.26	1.87	1.99	2.18	2.39	1.55	1.47	1.68	1.86
40. Dourado	1.39	1.66	1.53	2.68	1.38	1.96	1.94	1.58	2.58	1.35	1.55	1.51	1.62
41. Dourado Seco	0.99	1.21	0.88	2.31	0.42	1.24	1.95	1.88	2.48	1.68	1.87	1.53	1.52
42. Matao Branco	1.11	1.39	1.26	2.62	0.89	1.68	1.87	1.71	2.33	1.41	1.22	1.44	1.46
43. Pratao de Caspinas	1.38	1.67	1.54	2.89	1.67	1.92	1.86	1.49	2.46	1.48	1.57	1.47	1.52
44. Eorano	1.21	1.56	1.87	2.14	1.17	1.47	1.89	1.12	2.28	1.52	1.23	1.26	1.87

A respeito das distâncias entre pares de variedades, pode-se observar que suas magnitudes parecem ser, em termos gerais, crescentes em relação à genealogia, procedência geográfica e nível de melhoramento. Parentes próximos têm menores distâncias. Assim, três irmãos germanos (IAC 1246, IAC 5032 e IAC 5544) têm uma distância média $d = 1,31$; quatro meios irmãos (Dourado Agulha, IAC 1246, IAC 5032, IAC 5544) têm distância média $d = 1,61$. Entretanto, a relação avó-neto, que deveria ter valores similares ao de meios irmãos, por ter o mesmo coeficiente de parentesco ($F = 1/8$), resulta em um valor relativamente baixo ($d = 1,32$), mais semelhante à média de irmãos germanos ($F = 1/4$). Por isso, esta relação merece ser motivo de estudos detalhados para afirmações mais sólidas.

No que tange à relação distância e procedência geográfica, também em termos gerais, materiais de um mesmo Estado têm menor distância que quando confrontado variedades de Estados diferentes. Assim, a distância entre Amarelão IAMG N.14 e Jaguari ESAV IAMG N.25, ambas de Minas Gerais (MG), é $d = 0,94$; ao contrário de IAC 1246 de São Paulo (SP) e de Piemonte IAMG N.4 (MG) que estão distanciadas em $d = 2,55$. Tanto as relações entre distância e nível de melhoramento como a distância, serão abordadas com mais detalhes em parágrafos subsequentes.

As relações de afinidade entre as 58 variedades brasileiras, para fins de classificação, são melhor expressadas em agrupamentos. Pelo método de agrupamento de

Tocher (RAO, 1952), dez grupos de diversidade foram formados (Tabela 8). O grupo I inclui um grande número de variedades (42), os grupos 2 e 3 compreendem três variedades, os grupos 4, 5 e 6 abrangem duas variedades e os quatro grupos restantes foram singulares.

Tabela 8 - Composição de grupos de 58 variedades brasileiras de arroz pelo método de agrupamento de Tocher. Piracicaba, 1990.

GRUPO	NÚMERO DE VARIEDADES	COMPOSIÇÃO
I	42	Dourado Seco, Santa América, IAC 416, Biri-guí, Montanha Liso, Procópio, Ourinho, Prató, Prató Precoces, Batataes, Matão Branco, Fernandez, Agulhão, Iguape, Guaira, Bico Ganga, IAC 5032, Iguapão, 4 Meses, Cateto XXII, Dourado Liso, Goiano, Dourado Agulha, Jaguari ESAV N.25, Pindorama, Dourado Peludo, Dourado Precoces, Prató de Campinas, Dourado, Paraíba, Zebu, Amarelão, IAC 1246, Bacaba, IAC 47, 4 Meses n XL, Batataes, Caqui, Dourado Agulha 2, Agulha Branca, Pérola de Campinas, IAC 1391.
II	3	Prató de Uberlândia, Amarelão IAMG N.14, Amarelão IAMG N.2
III	3	IAC 25, Bico Preto, IAC 120
IV	2	Itarabe, Prató IAMG N.11
V	2	Guedes, Catetão
VI	2	Matão, Patnai 6
VII	1	Cateto Dourado
VIII	1	IAC 5544
IX	1	Agulha ESAV
X	1	Piemonte IAMG N.4

Além das limitações de informações sobre as variedades em questão, alguns aspectos interessantes podem ser observados deste agrupamento. Foram classificados nos grupos II a X, em sua maioria, materiais com algum tipo de melhoramento e de procedência diversa. Assim, tem-se variedades que vêm de MG (Amarelão IAMG N.2, Pratão IAMG N.11, Agulha ESAV, Piemonte IAMG N.4) e até introduções como o Patnai-6. Entretanto, uma parte dos materiais melhorados do IAC de SP, também foi agrupada nos grupos III (IAC 25 e IAC 120) e VIII (IAC 5544).

O grupo I é bastante numeroso, mas tem suficiente grau de semelhança que permite a sua aglomeração. Entretanto, envolve materiais dos Estados de São Paulo, Goiás, Maranhão e Minas Gerais e até variedades melhoradas e "não melhoradas". Algumas explicações a este fato são dados a seguir. As variedades melhoradas deste grupo, em sua maioria, foram obtidas através de seleção massal de variedades tradicionais, não sendo, por isso, muito diferenciadas de suas variedades originais e de outras variedades tradicionais. Assim, temos o caso de Dourado Seco, Pratão Precoces, Batalais e Bico Ganga. Entretanto, levando-se em conta que os materiais dentro do grupo estão, em média, em ordem crescente de diferenciação, outras observações podem ser consideradas. Os materiais melhorados de outros Estados resultam mais distanciados, como Bacaba (MA), Fernandez (GO), Jaguari ESAV IAMG N.25 (MG) e Zebu (MA).

Com relação a certos materiais melhorados do IAC, obtidos através de cruzamentos, algumas particularidades podem ser observadas. A IAC 25 ocupa uma distância tal que pode ser explicada pela distância de seus parentais, de acordo com a genealogia. Assim, o pai de IAC 25, IAC 1246, ocupa uma posição intermediária de distância. E se é comparada a distância de IAC 25, de seus parentais IAC 1246 ($d = 0,62$) e Dourado Precoce ($d = 1,61$), é reparado que está mais próximo de IAC 1246 que de Dourado Precoce. Assim, parece razoável supor que IAC 25 foi selecionado, levando-se em conta maiores características do IAC 1246. Inferências semelhantes podem ser obtidas da genealogia do IAC 47.

Nota: A distância $d = 1,23$ é a do IAC 47 a Prató; a distância do IAC 47 com Pérola não se tem informação.

No entanto, as relações entre coeficiente de parentesco e distância, calculadas a partir de dados eletroforéticos de semente, não podem ser demonstradas estatisticamente neste trabalho, devido ao número insuficiente de dados de parentesco, que permitam obter correlações significantes a respeito.

No agrupamento das variedades brasileiras pelo método do vizinho mais próximo, foram reiteradas as semelhanças e agrupamentos já notados no método anterior. Contudo, algumas precisões a mais foram obtidas. A aplicação deste método resulta em um dendograma (Figura 3) que é bem explicativo das relações de afinidade intra e intergrupos.

D I S T Â N C I A

Figura 3 - Dendrograma representativo entre 58 variedades brasileiras de arroz, obtido pelo método do vizinho mais próximo.

Pela análise do dendograma, pode-se verificar que as variedades Dourado Seco e Santa América são as mais próximas. Assim mesmo, o grande grupo I do método de Tocher mostra a sua variabilidade interna. O dendograma é demonstrativo que materiais de outros Estados diferentes a São Paulo estão mais diferenciados, como Bacava (MA), Fernandez (GO), Zebu (MA), dentro do citado grupo I. Classificando-se o grau de diferenciação entre todas as variedades, vistos no dendograma como baixo, médio e alto, tem-se geralmente, que materiais de grau médio são de MG, como Jaguari ESAV IAMG n 25, Amarelão IAMG nº 14 e Amarelão. Também estão no grau médio de diferenciação materiais de PB que são de irrigado: Caqui e Agulha Branca, além de variedades melhoradas do IAC: IAC 47, IAC 120 (irrigado) e IAC 5544. As variedades com alta diferenciação podem ser consideradas a partir de Catetão e inclui Cateto Dourado, Itarabe, Matão, Piemonte IAMG N. 4 e Agulha ESAV.

Por outro lado, algumas relações de parentesco são também expressas graficamente no dendograma. Assim, tem-se IAC 25 e IAC 1246, agrupadas proximamente.

Fazendo-se um apanhado geral dos dois métodos de agrupamento utilizados, pode-se dizer que seus resultados são semelhantes entre si. As relações de distância com a genealogia, procedência geográfica e estado de melhoramento, referidas na análise das distâncias, são novamente revelados nos agrupamentos.

No método de Tocher, faz-se uma classificação mais rigorosa. Assim o primeiro grupo fica com 42 variedades, cujo nível de semelhança é tal que pode se isolar dos demais. Os grupos de II a X classificam tanto materiais melhorados do Estado de São Paulo como de outros Estados, embora para estes últimos não sejam melhorados. Isto porque, pelo que parece, as pressões de seleção, seja natural ou artificial (melhoramento), ou ambas, ocasionam diferenças na composição genética da semente, que são reveladas na separação de proteínas, através de eletroforese de géis de poliacrilamida (SDS-PAGE).

Alguns resultados de agrupamentos pelo método de Tocher, aplicado sobre características morfológicas em arroz, são dados a seguir: VAIRAVAN (1973) sobre 190 coleções da região nordeste da Índia, encontrou nove grupos; MAURYA & SINGH (1977), a partir de 43 variedades, formaram 16 grupos; RAO *et alii* (1981), de 120 variedades, conseguiram 32 grupos e NADAF *et alii* (1986) obtiveram somente cinco grupos em 120 linhas elite.

Uma visão mais quantitativa do grau de diferenciação de variedades já referida é observado no dendograma da Figura 3. Nota-se um gradiente de diferenciação, onde os materiais mais diferentes (com maior valor no eixo das distâncias), correspondem a materiais melhorados e de outros Estados, excetuando-se os de São Paulo. Ainda, uma precisão a mais é relevada. Materiais de GO, MA e SP ocupam uma primeira porção desse gradiente. Já materiais de MG, PB

e SP ocuparam posições intermediárias. A porção terminal de maior nível de diferenciação inclui também materiais de MG e SP, muito embora, as reflexões sobre o gradiente de dispersão por Estado requeira estudos detalhados a esse respeito, para qualquer conclusão.

4.3. Diferenciação com análise de função discriminante (AFD)

O poder discriminante das avaliações de proteínas de grão de arroz foi testado através da AFD, inicialmente, entre todas as variedades (Figuras 4a, 4b e 4c). Pela natureza dos dados, os primeiros componentes não acumularam suficiente variabilidade (34,4% nos dois primeiros componentes) para explicar convenientemente, em gráficos de dispersão, o grau de afinidade entre todas as variedades. Estes valores não estão em concordância com aqueles obtidos pela maioria dos pesquisadores que aplicam AFD sobre dados morfológicos. Assim, na AFD de 23 variedades de *Vigna mungo* L., os dois primeiros componentes acumularam 70% de variabilidade total (DAS & GUPTA, 1984). RAO *et alii* (1981), ao analisar 120 variedades de arroz, acharam que os dois primeiros componentes aglomeraram 91% de diversidade. Mais recentemente, ERSKINE *et alii* (1989) descobriram que os três primeiros componentes reuniram 80% de variação em lentilha.

Figura 4 - Dispersão de escores discriminantes de 58 variedades brasileiras de arroz: a) escores do 1º componente; b) escores do 2º componente; c) escores do 1º e 2º componentes.

Entretanto, embora das limitações explicativas dos gráficos de dispersão as Figuras 4a, 4b e 4c são de proveito, pois as variedades mais diferenciadas conseguem se isolar. Assim, no gráfico do escore 1 (Figura 4a), as variedades 31 (Aguilha ESAV), 47 (Dourado Aguilha 2) e 53 (Amarelão IAMG N.2) conseguem se separar. Também as 14 (IAC 120) e 15 (IAC 47) juntas segregam-se das outras variedades. Entretanto, o gráfico do escore 2 (Figura 4b) é mais explicativo com relação aos agrupamentos de Tocher e de vizinho mais próximo. Neste grupo, foram isolados os grupos II (variedades 52, 53, 55), IV (49, 51), VIII (11), IX (31) e X (54). O grupo VII (38) foi parcialmente isolado e as variedades 12 e 14 do grupo III ocuparam lugares distantes bem diferenciados. O gráfico bidimensional dos escores 1 e 2 (Figura 4c) é também medianamente explicativo, e os grupos II, IV, V, IX e X podem ser diferenciados.

Segundo VAIRAVAN *et alii* (1973), as magnitudes absolutas dos coeficientes dos componentes da AFD refletem com boa extensão a importância dos caracteres para a diferenciação primária e secundária das variedades (neste caso). Entretanto, como foi dito, os dois primeiros componentes só acumularam 34,4% de variância total, por isso pode-se dizer que não há uma clara determinação de variáveis que contribuam substancialmente na diferenciação. Isto revela que a distinção entre as variedades evidencia-se mais claramente na somatória de diferenças relativamente pequenas de todas as características estudadas. Mesmo assim, pode-se afirmar,

do estudo dos componentes (Tabela 9) que, levando-se em conta as limitações anotadas, as variáveis de maior contribuição relativa (com os mais altos coeficientes no primeiro componente) foram as correspondentes às bandas 5 e 14 (ambas positivas). No segundo componente, destacam-se os coeficientes das bandas 12 e 8 (ambas de sinal negativo). No terceiro componente, são importantes os coeficientes das variáveis 12 e 4 (ambos com sinais positivos).

Tabela 9 - As primeiras 4 componentes da análise discriminante para a classificação de 58 variedades de arroz. Piracicaba, 1990.

COMPONENTE	1	2	3	4
Autovalor	4,8352	4,0042	3,3936	3,0753
Porcentagem de variância (%)	18,8	15,6	13,2	12,0

COEFICIENTES DOS COMPONENTES

Banda 1	0,1407	0,1799	0,3265	-0,0096
Banda 2	0,3310	-0,0341	0,0949	-0,1310
Banda 3	-0,1983	-0,2522	0,0649	-0,1948
Banda 4	0,1669	-0,3610	0,4465	0,2577
Banda 5	0,5955	-0,1358	-0,0067	-0,4646
Banda 6	-0,0135	-0,0579	-0,0344	0,3061
Banda 7	0,2012	-0,3022	0,3308	-0,0855
Banda 8	0,2204	-0,4437	0,0430	-0,1179
Banda 9	-0,0053	0,0407	0,1208	0,1016
Banda 10	0,1472	-0,2503	0,1777	0,1806
Banda 11	0,0172	-0,1318	0,1496	-0,3477
Banda 12	-0,0204	-0,4743	0,4933	0,0391
Banda 13	-0,2308	-0,1334	0,2459	0,3983
Banda 14	0,4422	-0,0304	0,1824	-0,3394
Banda 15	0,1739	-0,3334	0,2958	-0,2944
Banda 16	0,2414	-0,1611	0,2741	-0,1469

Uma utilidade maior do AFD no exame da capacidade discriminatória das variáveis em estudo foi a elaboração de funções discriminantes para grupos pré-estabelecidos.

4.3.1. Discriminação entre variedades melhoradas e "não melhoradas"

Aqui será feita análise, se é possível discriminar, sobre a base de avaliações quantitativas de frações protéicas em sementes de arroz, os grupos de variedades melhoradas e "não melhoradas". Foram considerados materiais melhorados, aquelas variedades que tiveram pelo menos algum processo de seleção por instituições oficiais (Tabela 1).

Os grupos de variedades melhoradas e "não melhoradas", tomados em conjunto, foram usados como base para uma AFD. Neste caso, o número mínimo de componentes é igual a um e a função discriminante correspondente é igual a:

$$Z = 0,3630 B_1 + 0,1537 B_2 - 0,1969 B_3 + 0,2667 B_4 + 0,4870 B_5 - \\ 0,2318 B_6 + 0,2379 B_7 - 0,0367 B_8 - 0,0122 B_9 - 0,0214 B_{10} + \\ 0,2472 B_{11} + 0,0253 B_{12} - 0,0200 B_{13} + 0,3527 B_{14} + 0,2801 B_{15} + \\ 0,2836 B_{16}$$

Esta função discriminante por ter um único componente aglomera 100% de variabilidade. O resultado do teste de Hotelling confirmou que existem diferenças significativas entre os valores médios para os dois grupos ($n_1 =$

16; $n_2 = 52$; $F(\text{obs}) = 2,83$ e nível de significância 0,2%. A diferença significativa entre o grupo de genótipos melhorados e "não melhorados" sugere que as pressões artificiais de seleção modificam a composição das frações protéicas da semente de arroz.

Com referência à importância das variáveis na discriminação destes dois grupos, pode-se observar nos coeficientes na função Z que as bandas 5, 1 e 14 têm as maiores grandezas. Mas, o mérito destas variáveis não pode também ser exagerado, pois seus valores não são tão marcadamente superiores com relação à média dos demais. Aqui também é válido dizer que a diferenciação entre variedades melhoradas e "não melhoradas" depende das contribuições de muitas variáveis, ressaltando, de alguma maneira, aquelas bandas assinaladas. Isto é relevado claramente quando são analisados a significância estatística, para o contraste em questão, e os gráficos de dispersão para um número crescente de variáveis. Assim, na análise univariada dos três variáveis com os maiores coeficientes na AFD, só o variável 1 é significativo ($F(\text{obs}) = 9,21$ ao 3% de probabilidade): No teste de significância para dois variáveis, através de uma análise multivariada de variância (MANOVA), foi encontrado que a combinação dos variáveis 1 e 5; e 1 e 14 não foi significativo; ao contrário os variáveis 1 e 14 foram altamente significativos. A MANOVA para os três variáveis (1, 5 e 14) foi também significativa ao 3% de probabilidade. Em relação aos níveis de diferenciação grupal (melhoradas e "não melhoradas"), mos-

trados através de gráficos de dispersão, os grupos formados tornam-se mais nítidos, quanto maior for o número de variáveis envolvidas na AFD sendo a mais apropriada aquela assinalada linhas atrás (com 16 variáveis) e que deu origem a Figura 5. A determinação de bandas com insignificante contribuição é mais clara. São de ajuda praticamente nula na diferenciação de variedades melhoradas e "não melhoradas" as bandas 8, 9, 10, 12 e 13.

Pelo número de grupos contrastados (2), melhorados e "não melhorados" um componente só foi obtido e, que alcançou 100% de variabilidade. Por isso, os gráficos de dispersão são bem explicativos do nível de diferenciação, tanto é que é possível obter estes dois grupos (Figura 5). Pode-se notar que, em termos gerais, as variedades melhoradas (MD) ocupam posições sobressalentes e as não melhoradas (NO) ficam relegadas a um plano inferior. Se o gráfico é dividido em faixas, que juntem o maior número de variedades do mesmo tipo, a delimitação entre grupos resulta um pouco mais clara: na faixa superior, dispõem-se, sobretudo, as variedades melhoradas (78% delas); na faixa inferior colocam-se 76% das não melhoradas; e, entre estas duas faixas, encontra-se uma zona de transição onde estão quase que em igual medida variedades melhoradas e "não melhoradas".

Figura 5 - Dispersão de escores discriminantes de 58 variedades brasileiras de arroz para o contraste de variedades melhoradas e "não melhoradas".

Entretanto a delimitação destas faixas não é absoluta, devido à existência de variedades melhoradas na faixa de materiais não melhorados. Isto pode ser explicado como segue: em termos gerais, pode-se dizer que as variedades melhoradas e não melhoradas não estão tão bem diferenciadas, que permitam um gráfico de dispersão nítido. E que existe também um gradiente de diferenciação na condição de melhorado. Este gradiente é esperado, sendo o melhoramento um processo gradativo de aprimoramento. Dos materiais melho-

rados que caíram na faixa dos não melhorados [Dourado Precoce (3), Amarelão (6), Pratão Precoce (7), Patnai 6 (32), Cateto XXII (37) e Dourado (4)], o pouco que se sabe é que na sua maioria são variedades melhoradas, mas obsoletas. Sabe-se, por exemplo, que Dourado e Dourado Precoce foram obtidas por seleção massal há muitos anos. Sabe-se também que o Pratão Precoce surgiu de uma mutação de Dourado Precoce. O Patnai 6 é um material introduzido que, ao que parece, não se adaptou adequadamente, pois num teste de tolerância à toxidez de alumínio, mostrou alta sensibilidade, diferindo da maioria das variedades melhoradas do IAC, (MODA-CIRINO, 1984).

Explicar a presença de quatro materiais não melhorados junto com materiais melhorados (Figura 5) não é uma tarefa fácil. O que é claro é que três delas estão bem próximas da faixa de transição; não é o mesmo caso do Pratão de Uberlândia (52) que está bem diferenciado. Outro aspecto importante a salientar é o destaque do IAC 47 (15) de reconhecido nível de melhoramento e do Piemonte IAMG N.4, variedade mineira da qual não se tem maiores informações. Mas, a variedade mais afastada foi o IAC 120 (14), uma variedade melhorada para irrigado, cuja posição foi contrastante com os outros materiais irrigados, como o Amarelão (6) e Caqui (24).

Com relação à importância dos variáveis analisados na diferenciação destes dois grupos, também resultou ser a somatória de pequenos subsídios de várias caracterís-

ticas, embora ocorra em menor escala que a diferenciação entre todas as variedades (item 4.3). E, diz-se em menor escala, porque sobre uma componente que armazena 100% de variabilidade total, as bandas 5, 1 e 14 têm peso relativamente maiores na diferenciação, em contraste com as bandas 9, 10 e 8, cujo auxílio é praticamente nulo.

A magnitude da heterose, que é o indicador mais seguro da divergência entre parentais (HALLAUER & MIRANDA, 1981), foi associado com certos padrões de esterase e peroxidase em arroz, (PENG *et alii*, 1988). Entretanto, o estudo destes últimos não mostrou nenhuma associação entre a magnitude da heterose na F1 e a variação enzimática entre parentais. Tal conclusão não foi surpreendente, segundo eles, considerando o pequeno número de marcadores usado no seu estudo, e a probabilidade de a heterose não ser uma função discreta de uns poucos componentes herdados. O que foi dito acima é para relacionar, de alguma maneira, o fato de não se encontrar determinadas bandas altamente relacionadas com materiais melhorados.

4.3.2. Discriminação entre variedades brasileiras e japonesas

O poder discriminante das frações proteicas de grãos de arroz também foi testado na separação entre variedades de duas origens distintas: brasileira e japonesa. O teste de significância de diferenças entre médias destes

dois grupos foi altamente significativo para o teste de Hotelling ($n_1 = 16$; $n_2 = 52$; $F(\text{obs}) = 20,63$). Isto pode ser enxergado, em forma manifesta, na Figura 6, onde as variedades japonesas ocupam posições superiores, diferenciando-se, desse modo, das brasileiras. Isto reitera a utilidade de avaliação das mencionadas frações protéicas, quando se deseja avaliar divergência genética.

Figura 6 - Dispersão de escores discriminantes de 67 variedades de arroz: 58 variedades brasileiras e 9 japonesas.

O estudo do poder discriminante das frações proteicas de grãos de arroz na distinção entre variedades brasileiras e japonesas foi bem mais claro que nos agrupamentos anteriores. Vê-se nitidamente que as variedades japonesas tiveram os maiores escores discriminantes, agrupando-se na parte superior do gráfico. Mais uma vez a AFD foi útil para verificar que existem diferenças médias significativas, nas variáveis analisadas com respeito às origens. Entretanto, como foi referido na revisão de literatura, a distância geográfica não é um indicador infalível de distância genética, e, fatores como seleção e deriva genética podem diferenciar bem mais que a origem geográfica. Mas, neste estudo, foi constatada uma diferenciação clara para os materiais japoneses; isto talvez esteja relacionado ao fato de ser o grão de variedades japonesas bem diferentes, em características morfológicas e bioquímicas, dos grãos de variedades brasileiras.

Entretanto, as relações de divergência entre materiais brasileiros e japoneses, detectadas através dos variáveis estudados, parecem estar relacionadas efetivamente a uma diversidade de alelos. CARBONERA (1990) encontrou manifestação positiva de heterose, heterobelitose e heterose padrão em 23 híbridos entre variedades, destes dois grupos. Para produção de grãos por planta, caráter mais importante avaliado, a heterose variou de -56,16% a 93,35%. Todavia a correlação da distância de Mahalanobis (calculado com dados morfológicos) com a manifestação de heterose nos híbridos

para produção de grãos foi positiva, mas baixa ($r = 0,18$) e não significativa. Contudo, são necessários estudos mais aprofundados para correlacionar distâncias multivariadas dos parentais, calculadas a partir de avaliações de frações proteicas, com a resposta heterótica.

A função discriminante para a diferenciação entre variedades brasileiras e japonesas foi:

$$Z = 0,2667 B_1 - 0,4119 B_2 + 0,3201 B_3 + 0,1458 B_4 - 0,5780 B_5 + 0,1481 B_6 - 0,1241 B_7 - 0,1073 B_8 + 0,1894 B_9 - 0,0598 B_{10} - 0,0201 B_{11} + 0,0947 B_{12} + 0,3714 B_{13} + 0,2223 B_{14} - 0,1140 B_{15} + 0,0567 B_{16}$$

Com relação à importância das variáveis na diferenciação dos grupos brasileiros e japoneses, nota-se na função discriminante, que existem variáveis que contribuem em forma um pouco mais importante na discriminação. Assim, as bandas 5, 2 e 13 têm, relativamente, altos coeficientes, mas que são insuficientes, quando utilizadas individualmente, na diferenciação clara destes grupos, e assim como referido na discriminação entre variedades melhoradas e "não melhoradas" a melhor separação resulta da função discriminante com a contribuição de todas as 16 variáveis, com certo destaque de aquelas mencionadas. Uma provável explicação a respeito seria que as pressões seletivas diferentes destes dois ambientes influenciam todas as frações proteicas analisadas, sendo maior em certo modo e nos grupos de moléculas que se originam e se manifestam nas bandas 5 e 2 na semente.

5. CONCLUSÕES

Os resultados obtidos do presente trabalho permitem apresentar as seguintes conclusões:

1. A avaliação quantitativa de frações proteicas de cariopse de arroz, por meio de SDS-PAGE, fornece suficiente variabilidade para o cálculo de distâncias multivariadas, com fins de classificação de variedades em grupos de diversidade.
2. O método de agrupamento de Tocher permitiu encontrar dez grupos divergentes, que podem ser úteis em programas de cruzamentos. Os grupos mais divergentes continham variedades de MG, preferencialmente. O dendograma do método do vizinho mais próximo, foi congruente com o método de Tocher e foi muito útil na descrição gráfica dos grupos de diversidade.
3. As avaliações das frações proteicas de grãos foram capazes de diferenciar em certo nível, através da AFD, variedades melhoradas e não melhoradas, revelando que os processos de melhoramento refletem-se nas frações

proteicas dos grãos, e podem ser reveladas por SDS-PAGE.

4. A diferenciação entre variedades de origem brasileira e japonesa é também possível, e, com maior nitidês através da AFD das referidas frações proteicas.
5. A diferenciação de variedades nos grupos testados com AFD (melhoradas e não melhoradas e brasileiras e japonesas) é, fundamentalmente, baseada no produto de somatória de pequenas diferenças em muitas variáveis. Não existem variáveis com pesos superlativos na discriminação.

REFERENCIAS BIBLIOGRÁFICAS

- ALIAGA-MOREL, J.R.; CULIANÉZ-MACIA, F.A.; CLEMENTE-MARIN, G.; PRIMO-YUFERA, E. Differentiation of rice varieties by eletrophoresis of embryo protein. Theoretical and Applied Genetics, Berlin, 74(2): 224-32, 1987.
- ARUNACHALAM, V. Genetic distance in plant breeding. Indian Journal of Genetics and Plant Breeding, New Delhi, 41(2): 226-36, 1981.
- BRADFORD, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, San Diego, 72(1-2): 248-54, 1976.
- BUSHUK, W. & ZILLMAN, R.R. Wheat cultivar identification by gliadin eletrophoregrams. I. Apparatus, method and nomenclature. Canadian Journal of Plant Science, Ottawa, 58: 505-15, 1978.
- CARBONERA, R. Avaliação da heterose e divergência genética em genótipos de arroz de sequeiro (*Oryza sativa* L.). Piracicaba, 1990. 103p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- CARRARO, D.M. Variação e herança dos padrões eletroforéticos em órgãos e estágios de desenvolvimento em milho (*Zea mays* L.). Piracicaba, 1990. 156p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).

- COOPER, T.G. The tools of Biochemistry. John Wiley e Sons, New York, 1977, 423p.
- CRUZ, C.D. Aplicação de algumas técnicas multivariadas no melhoramento de plantas. Piracicaba, 1990. 188p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- CURY, P.R. Análise de agrupamento: métodos sequenciais, aglomerativos e hierárquicos. Ciência e Cultura, São Paulo, 35(10): 1416-29, 1983.
- DAMERVAL, C.; HEBERT, Y.; de VIENNE, D. Is the polymorphism of protein amount related to phenotypic variability? Theoretical and Applied Genetics, 74: 194-202, 1987.
- DAS, P.K. & GUPTA, T.D. Multivariate analysis in black gram (*Vigna mungo* L. Hpper). Indian Journal of Genetics and Plant Breeding, New Delhi, 44(2): 243-7, 1984.
- DEMÉTRIO, C.G.B. Análise multidimensional para dados de cana-de-açúcar. Piracicaba, 1985. 144p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- DERBYSHIRE, E.; MÜLLER, H.P.; CARVALHO, M.T.V.; CROCOMO, O.J. Protein profiles of brazilian beans (*Phaseolus vulgaris*) obtained by electrophoresis in slabs of polyacrilamide gel. Energia Nuclear na Agricultura, Piracicaba, 3: 100-109, 1981.
- ERSKINE, W.; ADHAM, Y.; HOLLY, L. Geographic distribution of variation in quantitative traits in a world lentil collection. Euphytica, Wageningen, 43(1): 47-103, 1989.
- FONSECA, J.R. & BEDENDO, I.P. Características morfológicas, agrônômicas e fenológicas de algumas cultivares de arroz.

Goiânia, EMBRAPA/CNPAF, 1984. 57p. (CNPAF. Boletim de Pesquisa, 3).

FONSECA, J.R.; RANGEL, P.H.N.; PRABHU, A.S. Características botânicas e agronômicas de cultivares de arroz (*Oryza sativa* L.). Goiânia, EMBRAPA/CNPAF, 1981. 32p. (CNPAF. Circular Técnica, 14).

FORDE, B.G.; OLIVER, R.J.C.; LEAVER, C.J.; GUNN, R.E.; KEMBLE, R.J. Classification of normal and male-sterile cytoplasm in maize. I. Electrophoretic analysis of variation in mitochondrially synthesised proteins. *Genetics*, Princeton, 95(2): 443-50, 1980.

GHADERI, A.; ADAMS, M.W.; NASSIB, A.M. Relationship between genetic distance and heterosis for yield and morphological traits in dry edible beans and faba bean. *Crop Science*, Madison, 24(1): 37-42, 1984.

GODOY, C.R.M. *Análise estatística multidimensional*. Piracicaba, ESALQ, 1985. 187p.

GOODMAN, M.M. Distance analysis in biology. *Systematic Zoology*, Washington, 21(2): 174-86, 1972.

GUIMARÃES, R.C. Eletroforese em géis de poliacrilamida. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA PARA O PROGRESSO DA CIÊNCIA, 28., Brasília, 1976. Simpósios. São Paulo, SBPC, 1978. p.137-46.

HALLAUER, A.R. & MIRANDA FILHO, J.B. *Quantitative genetics in maize breeding*. Ames, State University Press, 1981. 468p.

HUSSAINI, S.H.; GOODMAN, M.M.; TIMOTHY, D.H. Multivariate analysis and the geographical distribution of the world

- collection of finger millet. Crop Science, Madison, 17: 257-63, 1977.
- JACQUARD, A. The genetic structure of populations. Berlin, Springer Verlag, 1974. 569p.
- JAIN, K.C.; PANDYA, B.P.; PANDE, K. Genetic divergence in chickpea. Indian Journal of Genetics and Plant Breeding, New Delhi, 41(2): 22-5, 1981.
- JOHNSON, B.L.; BARNHART, D.; HALL, O. Analysis of genome and species relationship in the polyploid wheats by protein electrophoresis. American Journal of Botany, New York, 54(9): 1089-98, 1967.
- LADIZINSKY, G. & HYMOWITZ, T. Seed protein electrophoresis in taxonomic and evolutionary studies. Theoretical and Applied Genetics, Berlin, 54(4): 145-51, 1979.
- LAEMMLI, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T₄. Nature, London, 227(5259): 680-5, 1970.
- LEFORT-BUSON, M. & VIENNE, D. Les distances génétiques. Paris, INRA Publications, 1985. 181p.
- McDANIEL, R.G. Electrophoretic characterization of proteins in *Hordeum*. Journal of Heredity, Washington, 61: 243-70, 1971.
- MAHALANOBIS, P.C. On the generalized distance in statistic. Proceedings of the National Institute of Sciences of India, New Delhi, 2(1): 49-55, 1936.
- MANLY, B.F.J. Multivariate statistical methods; a primer. New York, Chapman and Hall, 1986. 158p.

- MARCHYLO, B.A. & LA BERGE, D.E. Early cultivar identification by electrophoregram. Canadian Journal of Plant Science, Ottawa, 60(4): 1343-50, 1980.
- MAURYA, D.M. & SING, D.P. Genetic divergence in rice. Indian Journal of Genetics and Plant Breeding, New Delhi, 37(3): 395-402, 1977.
- MODA-CIRINO, V. Reação de genótipos de arroz (*Oryza sativa* L.) de sequeiro ao alumínio em solução nutritiva. Piracicaba, 1984. 162p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- MORISHIMA, H. & OKA, H. Phylogenetic differentiation of cultivated rice; XXII. Numerical evaluation of the Indica - Japonica differentiation. Japanese Journal of Breeding, Mishima, 31(4): 402-13, 1981.
- MURTY, B.R. & ARUNACHALAM, V. The nature of genetic divergence in relation to breeding system in crop plants. Indian Journal of Genetics and Plant Breeding, New Delhi, 26: 188-9, 1966.
- NADAF, S.K. & SING, I.S. Genetic divergence and multiple disease resistance studies in rice (*Oryza sativa* L.). International Rice Research Newsletter, Manila, 11(4): 16, 1986.
- OKA, H. & CHANG, W. Rice varieties intermediate between wild and cultivated forms and the origin of the Japonica type. Botanical Bulletin of Academia Sinica, Shanghai, 3(1): 109-31, 1962.
- OLIVEIRA, E.J. Análise multivariada no estudo da divergência genética entre cultivares de feijão (*Phaseolus vulgaris* L.). Viçosa, 1989. 91p. (Mestrado - Universidade Federal de Viçosa).

- PANDA, R.C.; KUMAR, O.A.; RAO, K.G.R. The use of protein electrophoresis in the study of phylogenetic relationships in Chili pepper (*Capsicum* L.). *Theoretical and Applied Genetics*, Berlin, 72(5): 665-70, 1986.
- PEETERS, J.P. & MARTINELLI, J.A. Hierarchical cluster analysis as a tool to manage variation in germplasm collection. *Theoretical and Applied Genetics*, Berlin, 78(1): 42-8, 1989.
- PENG, J.Y.; GUASZMANN, J.C.; VIRMANI, S.S. Heterosis and isozyme divergence in indica rice. *Crop Science*, Madison, 28(3): 561-3, 1988.
- PETER, K.V. & RAI, B. Genetic divergence in tomato. *Indian Journal of Genetics and Plant Breeding*, New Delhi, 36(3): 379-83, 1976.
- RAM, J. & PANWAR, D.V.S. Interspecific divergence in rice. *Indian Journal of Genetics and Plant Breeding*, New Delhi, 30(1): 1-10, 1970.
- RAO, C.R. Advanced statistical methods in biometrical research. John Wiley e Sons, New York, 1952. 390p.
- RAO, A.Y.; PRASAD, A.S.R.; SAI KRISHNA, T.; SESHU, D.V.; SRINIVASAM, T.E. Genetic divergence among some brown planthopper resistant rice varieties. *Indian Journal of Genetics and Plant Breeding*, New Delhi, 41(1/2): 135-45, 1989.
- RISSI, J.C. & GALWEY, N.W. The pattern of genetic diversity in the Andean grain crop quinoa (*Chenopodium quinoa* Willd); II. Multivariate methods. *Euphytica*, Wageningen, 41(1/2): 135-45, 1989.

- SARKAR, R. & BOSE, S. Electrophoretic characterization of rice variations using single seed (salt soluble) proteins. *Theoretical and Applied Genetics*, Berlin, 68(5): 415-9, 1984.
- SIDDIQ, E. A.; NERKAR, Y. S.; MEHTA, J. L. Intra and inter-specific variation in soluble proteins of *Oryza sativa* L.. *Theoretical and Applied Genetics*, Berlin, 42(8): 351-6, 1972.
- SINGH, R. K. & CHAUDHARY, B. D. Biometrical methods in quantitative genetic analysis, New Delhi, Kalyani Publishers, 1979. 304p.
- SMITH, J. S. C. & SMITH, O. S. The description and assesment of distance between inbred lines of maize; I. The use of morphological traits as descriptors. *Maydica*, Bergamo, 34(2): 141-50, 1989a.
- SMITH, J. S. C. & SMITH, O. S. The description and assesment of distance between inbred lines of maize; II. The utility of morphological, biochemical and genetic descriptors and scheme for the testing of distinctiveness between inbred lines. *Maydica*, Bergamo, 34(2): 151-61, 1989b.
- UNGAR, A. I. & BOUCAUD, J. Comparison of seed proteins in the genus *Suaeda* (Chenopodiaceae) by means of disc gel electrophoresis. *American Journal of Botany*, New York, 61(4): 325-30, 1974.
- VAIRAVAN, S.; SIDDIQ, E. A.; ARUNACHALAM, V.; SWAMINATHAM, M. S. A study on the genetic divergence in rice from North-East Himalaya. *Theoretical and Applied Genetics*, Berlin, 43(5): 213-21, 1973.

VALLS, J.F.M. Caracterização morfológica, reprodutiva e bioquímica de germoplasma vegetal. In: ENCONTRO SOBRE RECURSOS GENÉTICOS, 1., Jaboticabal, 1988. Anais. Jaboticabal, UNESP/FCAVJ, 1989. p.106-28.

VARMA, N.S. & GULATI, S.C. Genetic divergence in 2-rowed and 6-rowed barley. Indian Journal of Genetics and Plant Breeding, New Delhi, 42(3): 314-8, 1982.

WEN, T.N. & LUTHE, D.S. Biochemical characterization of rice glutelin. Plant Physiology, Lancaster, 78(1): 172-7, 1985.

WRIGHT, S. Evolution and the genetics of populations. Volume 4: Variability within among natural populations. University of Chicago Press, Chicago, 1978. 580p.